

UNIVERSIDADE FEDERAL DO PARANÁ

JOÃO FELIPE CARDOZO MARTINS

MÉTODOS DE APRENDIZAGEM DE MÁQUINA APLICADOS NA MODELAGEM DE  
BIOMASSA DE ÁRVORES USANDO A LINGUAGEM DE PROGRAMAÇÃO *JULIA*

CURITIBA

2020

JOÃO FELIPE CARDOZO MARTINS

MÉTODOS DE APRENDIZAGEM DE MÁQUINA APLICADOS NA MODELAGEM DE  
BIOMASSA DE ÁRVORES USANDO LINGUAGEM DE PROGRAMAÇÃO *JULIA*

Dissertação apresentada ao curso de Pós-Graduação em Engenharia Florestal, Setor de Ciências Agrárias, Universidade Federal do Paraná, como requisito parcial à obtenção do título de Mestre em Engenharia Florestal.

Orientador: Prof. Dr. Alexandre Behling

Coorientador: Prof. Dr. Allan Libanio Pelissari  
Prof. Dr. Richardson Ribeiro

CURITIBA

2020

Ficha catalográfica elaborada pela  
Biblioteca de Ciências Florestais e da Madeira - UFPR

Martins, João Felipe Cardozo

Métodos de aprendizagem de máquina aplicados na modelagem de biomassa de árvores usando a linguagem de programação Julia / João Felipe Cardozo Martins. - Curitiba, 2020.

143 f. : il.

Orientador: Prof. Dr. Alexandre Behling

Coorientadores: Prof. Dr. Allan Libanio Pelissari

Prof. Dr. Richardson Ribeiro

Dissertação (Mestrado) - Universidade Federal do Paraná, Setor de Ciências Agrárias, Programa de Pós-Graduação em Engenharia Florestal. Defesa: Curitiba, 21/02/2020.

Área de concentração: Manejo Florestal.

1. Biomassa florestal. 2. Julia (Linguagem de programação de computador). 3. Florestas – Métodos estatísticos. 4. Acácia. 5. Florestas – Medição. 6. Teses. I. Behling, Alexandre. II. Pelissari, Allan Libanio. III. Ribeiro, Richardson. IV. Universidade Federal do Paraná, Setor de Ciências Agrárias. V. Título.

CDD – 634.9

CDU – 634.0.548

Bibliotecária: Berenice Rodrigues Ferreira – CRB 9/1160

## TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em ENGENHARIA FLORESTAL da Universidade Federal do Paraná foram convocados para realizar a arguição da Dissertação de Mestrado de **JOÃO FELIPE CARDOZO MARTINS** intitulada: **MÉTODOS DE APRENDIZAGEM DE MÁQUINA APLICADOS NA MODELAGEM DE BIOMASSA DE ÁRVORES USANDO A LINGUAGEM DE PROGRAMAÇÃO JULIA**, sob orientação do Prof. Dr. ALEXANDRE BEHLING, que após terem inquirido o aluno e realizada a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa. A outorga do título de mestre está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

CURITIBA, 21 de Fevereiro de 2020.



ALEXANDRE BEHLING

Presidente da Banca Examinadora (UNIVERSIDADE FEDERAL DO PARANÁ)



FABRICIO ENEMBRECK

Avaliador Externo (PONTIFÍCIA UNIVERSIDADE CATÓLICA DO PARANÁ)



HENRIQUE SOARES KOEHLER

Avaliador Interno (UNIVERSIDADE FEDERAL DO PARANÁ)

A minha irmã Manuela e meu irmão Thomas Antônio, pelo tempo que deixamos de estar juntos e a todos os estresses causados por minha vida acadêmica.

Aos meus pais, João e Daezi, a eles todos os créditos por ter chegado aonde cheguei, sou o que sou, em grande parte, por eles.

A minha namorada Dniffer Martins por estar ao meu lado e ter paciência mesmo em um fim de ano sem poder contar com a minha presença em tempo integral

Dedico a eles essa dissertação.



## **AGRADECIMENTOS**

Universidade Federal do Paraná, Programa de Pós-graduação em Engenharia Florestal, pela oportunidade de estudo. Aos professores da pós-graduação, o meu obrigado pelos conhecimentos transmitidos.

À minha família, por alegrar minha vida. Especialmente minha mãe Deise, batalhadora que criou seus três filhos diante de muitas dificuldades. Esse espírito de luta me fez crescer e me fortalece em tudo que busco.

A minha equipe de trabalho, sua organização tornou meu desafio de lidar com o tempo possível.

Ao Prof. Dr. Alexandre Behling, pela dedicação nas correções e orientações neste período de aprendizado e pela amizade adquirida durante esse longo período de cooperação.

Ao Prof. Allan Libanio Pelissari, que mudou minha visão sobre a sala de aula e auxiliou na construção da presente tese, gerando grandes acréscimos para mesma.

Ao Professor Richardson Ribeiro que contribui com meu ingresso na área de programação e auxiliou na descoberta desse novo mundo.

Aos demais professores da Universidade Federal do Paraná que me propiciaram conhecimentos e foram determinantes em minha formação. Aos amigos de jornada, nos estudos, no laboratório e experimentos conduzidos, deixo aqui meu agradecimento especial.

Aos meus colegas de pós-graduação que tornaram um período de longa dedicação em algo divertido.

Aos meus amigos que permaneceram firmes ao meu lado durante esses anos, sendo consideradas pessoas muito Originais.

A empresa TANAGRO /TANAC S.A. pelos dados disponibilizados, sem eles não haveria como realizar os avanços científicos que este trabalho proporcionou.

Agradecimentos à Capes e ao CNPq pelo incentivo prestado à pesquisa ao disponibilizar bolsas de pós-graduação.

Aos secretários do Programa de Pós-graduação em Engenharia Florestal da Universidade Federal do Paraná por sempre terem dispensado preciosa atenção.

*"O melhor de mim é aquilo que desconheço.  
Aquilo que eu ainda não sei, aquilo que me coloca  
numa rota de criatividade e não de esgotamento."*

*Mario Sérgio Cortella.*

## RESUMO

O contexto deste trabalho visa modelar a biomassa total de árvores por meio de técnicas de aprendizagem de máquina utilizando a linguagem de programação *Julia*. A importância em quantificar a biomassa nas árvores está associada aos cenários de fixação de carbono, ciclagem de nutrientes, fins energéticos e também para avaliação do crescimento da floresta; contudo quantificar a biomassa de uma floresta é um desafio para os profissionais da área florestal, dada as dificuldades em mensurar essa variável. Usualmente, a biomassa é estimada por meio de equações de regressão, mas outras técnicas têm sido empregadas com objetivo de obter estimadores mais acurados, entre elas, técnicas de aprendizado de máquina. Nessa direção, foram propostos os seguintes objetivos: i) Aplicar e avaliar técnicas e algoritmos de aprendizagem de máquina: redes neurais artificiais, máquinas de vetores de suporte, florestas aleatórias e *k*-vizinhos mais próximos na modelagem de biomassa da espécie acácia-negra; ii) Comparar o melhor resultado de aprendizagem de máquina com equações de regressão, considerada como a técnica testemunha; e iii) realizar as modelagens utilizando a linguagem de programação *Julia*. Para tanto, foram utilizadas duas bases de dados da espécie acácia-negra, uma para treinamento composta por 670 árvores e outra para validação com 96 árvores. Na modelagem, a biomassa foi tratada como a variável dependente, sendo que o diâmetro à altura do peito, a altura total e a idade foram as variáveis independentes. Floresta aleatória foi a melhor técnica de aprendizagem de máquina para modelar a biomassa de árvores de acácia-negra. Essa técnica também apresentou maior capacidade para estimar a biomassa de árvores para novos dados, quando comparado com as equações de regressão, portanto, é recomendada. Outra contribuição desse trabalho foi explorar a linguagem de programação *Julia*, demonstrando seu potencial em problemas da Ciência Florestal.

Palavras-chave: ciência de dados; linguagens de programação; floresta aleatória; análise de regressão; acácia-negra



## ABSTRACT

The context of the present endeavor has the purpose of modeling Total Tree Biomass, by means of implementing Machine Learning Techniques, using the Programming Language referred to as, "Julia". The importance in quantifying Tree Biomass is associated with the following scenarios: carbon fixation, nutrient cycling, energy purposes and also for the evaluating of forest growth; however, achieving Forest Biomass Quantification is a challenge for professionals, who act in the realm of forestation, given the difficulties inherent in measuring this variable. Usually, Forest Biomass is estimated using regression related equations, but other techniques have come to the fore encumbering the minds of researchers, who have the goal of striving to obtain more accurate estimation techniques, among them being, Machine Learning techniques. In this regard, the following objectives have been proposed: i) implementing and evaluating techniques as well as Machine Learning algorithms: artificial neural networks, support vector machines, random forests and k-nearest neighbors, in the biomass modeling of the black acacia species; ii) Comparing Machine Learning best results with regression equations, which are considered as the baseline technique; and iii) Carrying out the modeling using the said Programming Language, Julia. Therefore; two databases have been used stemming from the black wattle species, one for training related purposes, consisting of 670 trees and another for validation purposes comprising 96 trees. In the modeling exercise, the Forest Biomass in question was treated as the dependent variable, while the Diameter at Breast Height (DBH), Total Height and Age were factored as the independent variables. The random forest was deemed as the best Machine Learning technique for the purposes of modeling the biomass of black wattle trees. The aforementioned technique also revealed a greater ability to ascertain the Tree Biomass for obtaining new data, when compared to regression related equation techniques and is; therefore, the recommended technique. Another contributing factor of this study was that it afforded exploration into the said Programming Language, Julia, by demonstrating the Former's potential for tackling the problems of Forest Science itself, having inherently high IT related costs, such as, in the case of implementing Machine Learning modeling techniques.

Keywords: date science; programming languages; random forest; regression related analysis; black wattle

## LISTA DE ALGORITMOS

ALGORITMO 1 – IMPORTAÇÃO DE BASE DE DADOS NA LINGUAGEM DE PROGRAMAÇÃO <i>JULIA</i> .....	55
ALGORITMO 2 – IMPORTAÇÃO DE <i>PACKAGES</i> NA LINGUAGEM DE PROGRAMAÇÃO <i>JULIA</i> .....	58
ALGORITMO 3 – CÁLCULO DA ESTATÍSTICA DESCRITIVA NA LINGUAGEM DE PROGRAMAÇÃO <i>JULIA</i> .....	64
ALGORITMO 4 – PROGRAMAÇÃO PARA CRIAÇÃO DOS GRÁFICOS: <i>VIOLINPLOT</i> , <i>BOXPLOT</i> E <i>DOTPLOT</i> NA LINGUAGEM DE PROGRAMAÇÃO <i>JULIA</i> .....	66
ALGORITMO 5 – REPRESENTAÇÃO DA PROGRAMAÇÃO DE <i>CORNERPLOT</i> EM LINGUAGEM DE PROGRAMAÇÃO <i>JULIA</i> . ....	71
ALGORITMO 6 – PROGRAMAÇÃO PARA CRIAÇÃO DOS GRÁFICOS DE DISPERSÃO PARA COMPARAÇÃO NA LINGUAGEM DE PROGRAMAÇÃO <i>JULIA</i> .....	75
ALGORITMO 7 – PROGRAMAÇÃO PARA CRIAÇÃO DOS GRÁFICO COM LINHA DE TENDÊNCIA DAS BASE DE DADOS 670/2014 E 98/2012 PARA COMPARAÇÃO NA LINGUAGEM DE PROGRAMAÇÃO <i>JULIA</i> .....	76
ALGORITMO 8 – REPRESENTAÇÃO DA PROGRAMAÇÃO DAS ESTÁTISTICAS VIÉS, PRECISÃO E ACURÁCIA, DENOMINADA FUNÇÃO VPA .....	78
ALGORITMO 9 – REPRESENTAÇÃO DA PROGRAMAÇÃO DOS GRÁFICOS DE DISPERSÃO PARA OS ESTIMADORES DE BIOMASSA EM LINGUAGEM DE PROGRAMAÇÃO <i>JULIA</i> . ....	79
ALGORITMO 10 – REPRESENTAÇÃO DA PROGRAMAÇÃO DOS GRÁFICOS DE RESÍDUOS PARA OS ESTIMADORES DE BIOMASSA EM LINGUAGEM DE PROGRAMAÇÃO <i>JULIA</i> . ....	80
ALGORITMO 11 – REPRESENTAÇÃO DA PROGRAMAÇÃO DE REDES NEURAIS ARTIFICIAIS PARA ESTIMAR A BIOMASSA EM LINGUAGEM DE PROGRAMAÇÃO <i>JULIA</i> .....	81
ALGORITMO 12 – REPRESENTAÇÃO DA PROGRAMAÇÃO DE TÉCNICAS DE APRENDIZAGEM DE MÁQUINA PARA ESTIMAR BIOMASSA DE DADOS INDEPENDENTES AOS DADOS DE AJUSTE EM LINGUAGEM DE PROGRAMAÇÃO <i>JULIA</i> . ....	86

ALGORITMO 13 – REPRESENTAÇÃO DA PROGRAMAÇÃO DE $k$ -VIZINHOS MAIS PRÓXIMOS PARA ESTIMAR A BIOMASSA EM LINGUAGEM DE PROGRAMAÇÃO <i>JULIA</i> .....	87
ALGORITMO 14 – REPRESENTAÇÃO DA PROGRAMAÇÃO DE FLORESTAS ALEATÓRIAS PARA ESTIMAR A BIOMASSA EM LINGUAGEM DE PROGRAMAÇÃO <i>JULIA</i> .....	92
ALGORITMO 15 – REPRESENTAÇÃO DA PROGRAMAÇÃO DE MÁQUINAS DE VETORES DE SUPORTE PARA ESTIMAR A BIOMASSA EM LINGUAGEM DE PROGRAMAÇÃO <i>JULIA</i> . ....	96
ALGORITMO 16 – REPRESENTAÇÃO DA PROGRAMAÇÃO DE EQUAÇÕES NÃO-LINEARES PARA ESTIMAR A BIOMASSA EM LINGUAGEM DE PROGRAMAÇÃO <i>JULIA</i> .....	108
ALGORITMO 17 – REPRESENTAÇÃO DA PROGRAMAÇÃO DE EQUAÇÕES NÃO-LINEARES PARA ESTIMAR A BIOMASSA DA BASE DE DADOS DE VALIDAÇÃO EM LINGUAGEM DE PROGRAMAÇÃO <i>JULIA</i> .....	111

## LISTA DE FIGURAS

FIGURA 1 – APRENDIZAGEM SUPERVISIONADA E NÃO-SUPERVISIONADA ...	33
FIGURA 2 – ESTRUTURA DE UM NEURÔNIO APRESENTANDO OS SEUS VALORES DE ENTRADA, SEUS PESOS, O SOMATÓRIO REALIZADO E A FUNÇÃO DE ATIVAÇÃO QUE RESULTA NA SUA SAÍDA. ....	35
FIGURA 3 – ARQUITETURA HIPOTÉTICA DE UMA REDE NEURAL ARTIFICIAL DE MÚLTIPLAS CAMADAS.....	36
FIGURA 4 – ESQUEMA DE FUNCIONAMENTO DE UMA FLORESTA ALEATÓRIA. A ENTRADA É APRESENTADA A TODAS AS ÁRVORES DA FLORESTA E O RESULTADO DA PREDIÇÃO É A MÉDIA DOS RESULTADOS DE CADA ÁRVORE.....	39
FIGURA 5 – ESQUEMA DE FUNCIONAMENTO DE UM ALGORITMO DE <i>k</i> -VIZINHOS MAIS PRÓXIMOS.....	41
FIGURA 6 – ILUSTRAÇÃO DE UM PROBLEMA DE CLASSIFICAÇÃO BINÁRIA LINEARMENTE SEPARAVEL, COM SOLUÇÕES POSSÍVEIS PARA O PROBLEMA DE SEPARAÇÃO DOS CONJUNTOS ATRAVÉS DE UM HIPERPLANO .....	44
FIGURA 7 – POVOAMENTO DE ACÁCIA-NEGRA COM IDADES DE 7 E 5 ANOS, RESPECTIVAMENTE, NO ESTADO DO RIO GRANDE DO SUL, BRASIL. ....	49
FIGURA 8 – ÁREA DE CULTIVO DE ACÁCIA-NEGRA NAS REGIÕES DE CRISTAL, ENCRUZILHADA DO SUL E PIRATINI, NO ESTADO DO RIO GRANDE DO SUL, BRASIL.....	51
FIGURA 9 – REPRESENTAÇÃO DA INTERFACE DA LINGUAGEM DE PROGRAMAÇÃO <i>JULIA</i> UTILIZANDO <i>REPL</i> . ....	54
FIGURA 10 – BASE DE DADOS 670/2014 CARREGADA NO AMBIENTE DA LINGUAGEM DE PROGRAMAÇÃO <i>JULIA</i> .....	55
FIGURA 11 – REDE NEURAL ARTIFICIAL REPRESENTATIVA PARA A FORMULAÇÃO UTILIZADA NA PREDIÇÃO DOS DADOS 670/2014 .....	59
FIGURA 12 – GRÁFICO <i>VIOLINPLOT</i> , <i>BOXPLOT</i> E <i>DOTPLOT</i> PARA A VARIÁVEL $d$ (cm), PARA AS CLASSES DE IDADES: JOVEM, MÉDIA INICIAL,	

MÉDIA AVANÇADA E MADURA PARA A BASE DE DADOS 670/2014. .....	67
FIGURA 13 – GRÁFICO <i>VIOLINPLOT</i> , <i>BOXPLOT</i> E <i>DOTPLOT</i> PARA A VARIÁVEL $d$ (cm), PARA AS CLASSES DE IDADES: JOVEM, MÉDIA INICIAL E MÉDIA AVANÇADA PARA A BASE DE DADOS 98/2012.....	67
FIGURA 14 – GRÁFICO <i>VIOLINPLOT</i> , <i>BOXPLOT</i> E <i>DOTPLOT</i> PARA A VARIÁVEL ALTURA (m) DE ÁRVORES DE ACÁCIA NEGRA, PARA AS CLASSES DE IDADES: JOVEM, MÉDIA INICIAL, MÉDIA AVANÇADA E MADURA PARA A BASE DE DADOS 670/2014. ....	68
FIGURA 15 – GRÁFICO <i>VIOLINPLOT</i> , <i>BOXPLOT</i> E <i>DOTPLOT</i> PARA A VARIÁVEL ALTURA (m) DE ÁRVORES DE ACÁCIA NEGRA, PARA AS CLASSES DE IDADES: JOVEM, MÉDIA INICIAL E MÉDIA AVANÇADA PARA A BASE DE DADOS 98/2012.....	69
FIGURA 16 – GRÁFICO <i>VIOLINPLOT</i> , <i>BOXPLOT</i> E <i>DOTPLOT</i> PARA A VARIÁVEL BIOMASSA DE ACÁCIA NEGRA (Kg), PARA AS CLASSES DE IDADES: JOVEM, MÉDIA INICIAL, MÉDIA AVANÇADA E MADURA PARA A BASE DE DADOS 670/2014.....	69
FIGURA 17 – GRÁFICO <i>VIOLINPLOT</i> , <i>BOXPLOT</i> E <i>DOTPLOT</i> PARA A VARIÁVEL BIOMASSA DE ACÁCIA NEGRA (Kg), PARA AS CLASSES DE IDADES: JOVEM, MÉDIA INICIAL E MÉDIA AVANÇADA PARA A BASE DE DADOS 98/2012.....	70
FIGURA 18 – GRÁFICO <i>CORNERPLOT</i> PARA AS VARIÁVEIS DIÂMETRO À ALTURA DO PEITO ( $d$ ), EM cm, ALTURA TOTAL ( $h$ ), EM m, BIOMASSA TOTAL ( $b$ ), EM Kg E IDADE ( $i$ ), EM anos, DE ACÁCIA NEGRA, PARA A BASE DE DADOS 670/2014. ....	71
FIGURA 19 – RELAÇÃO ALOMÉTRICA DA BIOMASSA DE ÁRVORES DE ACÁCIA-NEGRA COM AS VARIÁVEIS $d$ (cm) e $h$ (m) PARA A BASE DE DADOS 670/2014 E PARA A 98/2012.....	75
FIGURA 20 – LINHAS DE TENDÊNCIA DA DISPERSÃO DA BIOMASSA DE ÁRVORES DE ACÁCIA-NEGRA COM AS VARIÁVEIS $d$ (cm) e $h$ (m) PARA A BASE DE DADOS 670/2014 E PARA A 98/2012. ....	77
FIGURA 21 – BASE DE DADOS 670/2014 CONTENDO ESTIMADORES DE BIOMASSA POR ESTIMADORES GENÉRICOS E SEUS	

RESPECTIVOS RESULTADOS PARA CADA ÁRVORE DE ACÁCIA- NEGRA PARA EXEMPLIFICAR. ....	78
FIGURA 22 - VALORES OBSERVADOS DA VARIÁVEL BIOMASSA DE ÁRVORES DE ACÁCIA NEGRA E ESTIMADOS POR REDES NEURAI ARTIFICIAIS. ....	84
FIGURA 23 – GRÁFICO DE RESÍDUOS (em kg) DA VARIÁVEL BIOMASSA DE ÁRVORES DE ACÁCIA NEGRA MODELADA POR REDES NEURAI ARTIFICIAIS. ....	85
FIGURA 24 – VALORES OBSERVADOS DA VARIÁVEL BIOMASSA DE ÁRVORES DE ACÁCIA NEGRA E ESTIMADOS POR <i>k</i> -VIZINHOS MAIS PRÓXIMOS.....	89
FIGURA 25 – GRÁFICO DE RESÍDUOS (em kg) DA VARIÁVEL BIOMASSA DE ÁRVORES DE ACÁCIA NEGRA MODELADA POR <i>k</i> -VIZINHOS MAIS PRÓXIMOS.....	90
FIGURA 26 – VALORES OBSERVADOS DA VARIÁVEL BIOMASSA DE ÁRVORES DE ACÁCIA NEGRA E ESTIMADOS POR FLORESTAS ALEATÓRIAS.....	94
FIGURA 27 – GRÁFICO DE RESÍDUOS (em Kg) DA VARIÁVEL BIOMASSA DE ÁRVORES DE ACÁCIA NEGRA MODELADA POR FLORESTAS ALEATÓRIAS.....	94
FIGURA 28 – VALORES OBSERVADOS DA VARIÁVEL BIOMASSA DE ÁRVORES DE ACÁCIA NEGRA E ESTIMADOS POR MÁQUINAS DE VETORES DE SUPORTE.....	98
FIGURA 29 – GRÁFICO DE RESÍDUOS (em Kg) DA VARIÁVEL BIOMASSA DE ÁRVORES DE ACÁCIA NEGRA MODELADA POR MÁQUINAS DE VETORES DE SUPORTE.....	99
FIGURA 30 – GRÁFICO DE DISPERSÃO DOS ESTIMADORES POR MÁQUINA DE VETORES DE SUPORTE EM COMPARAÇÃO COM A BIOMASSA TOTAL. ....	102
FIGURA 31 – GRÁFICO DE RESÍDUOS DOS ESTIMADORES POR MÁQUINA DE VETORES DE SUPORTE EM COMPARAÇÃO COM A BIOMASSA TOTAL. ....	103



FIGURA 32 – VALORES OBSERVADOS E ESTIMADOS DA VARIÁVEL BIOMASSA DE ÁRVORES DE ACÁCIA NEGRA E RESÍDUOS (em kg) OBTIDOS POR MEIO DE MODELOS DE REGRESSÃO.....	110
---	-----

## LISTA DE TABELAS

TABELA 1 – CLASSES DE IDADE DOS POVOAMENTOS DE ACÁCIA-NEGRA AMOSTRADOS NO ESTADO DO RIO GRANDE DO SUL, BRASIL.	52
TABELA 2 – NÚMERO DE ÁRVORES AMOSTRADAS ( $n$ ) PARA A BASE DE DADOS 98/2012 DE ACÁCIA-NEGRA NO ESTADO DO RIO GRANDE DO SUL, BRASIL. ....	54
TABELA 3 – <i>PACKAGES</i> UTILIZADOS NA MODELAGEM DA BIOMASSA UTILIZANDO A LINGUAGEM DE PROGRAMAÇÃO <i>JULIA</i> .....	57
TABELA 4 – EQUAÇÕES DE BIOMASSA TOTAL ACIMA DO SOLO PARA ÁRVORES DE ACÁCIA-NEGRA ORIUNDAS DE POVOAMENTOS NO ESTADO DO RIO GRANDE DO SUL, BRASIL. ....	61
TABELA 5 – ESTATÍSTICA DESCRITIVA PARA AS VARIÁVEIS DIÂMETRO À ALTURA DO PEITO (cm), ALTURA TOTAL (m) E BIOMASSA TOTAL (kg) PARA A BASE DE DADOS DE AJUSTES,.....	65
TABELA 6 – ESTATÍSTICA DESCRITIVA PARA AS VARIÁVEIS DIÂMETRO À ALTURA DO PEITO (cm), ALTURA TOTAL (m) E BIOMASSA TOTAL (Kg) PARA A BASE DE DADOS DE VALIDAÇÃO 98/2012.....	66
TABELA 7 – RESULTADOS DE VIÉS, PRECISÃO E ACURÁCIA PARA OS ESTIMADORES DA VARIÁVEL BIOMASSA DE ÁRVORES DE ACÁCIA NEGRA MODELOS DE REDES NEURAIIS ARTIFICIAIS, UTILIZANDO-SE A BASE DE DADOS 670/2014. ....	84
TABELA 8 – RESULTADOS DE VIÉS, PRECISÃO E ACURÁCIA PARA OS ESTIMADORES DA VARIÁVEL BIOMASSA DE ÁRVORES DE ACÁCIA NEGRA UTILIZANDO MODELOS DE REDES NEURAIIS ARTIFICIAIS, PARA A BASE DE DADOS DE VALIDAÇÃO 98/2012. ....	87
TABELA 9 – RESULTADO DE VIÉS, PRECISÃO E ACURÁCIA PARA OS ESTIMADORES POR ALGORITMO DE $k$ -VIZINHOS MAIS PRÓXIMOS.....	89
TABELA 10 – RESULTADOS DE VIÉS, PRECISÃO E ACURÁCIA PARA OS ESTIMADORES DA VARIÁVEL BIOMASSA DE ÁRVORES DE ACÁCIA NEGRA UTILIZANDO MODELOS DE $k$ -VIZINHOS MAIS PRÓXIMOS, PARA A BASE DE DADOS DE VALIDAÇÃO 98/2012.	91

TABELA 11 – RESULTADOS DE VIÉS, PRECISÃO E ACURÁCIA PARA OS ESTIMADORES DA VARIÁVEL BIOMASSA DE ÁRVORES DE ACÁCIA NEGRA UTILIZANDO-SE MODELOS DE REDES NEURAIS ARTIFICIAIS, PARA A BASE DE DADOS DE AJUSTE 670/2014.....	93
TABELA 12 – RESULTADOS DE VIÉS, PRECISÃO E ACURÁCIA PARA OS ESTIMADORES DA VARIÁVEL BIOMASSA DE ÁRVORES DE ACÁCIA NEGRA UTILIZANDO MODELOS DE FLORESTAS ALEATÓRIAS, PARA A BASE DE DADOS DE VALIDAÇÃO 98/2012. .....	95
TABELA 13 – RESULTADOS DE VIÉS, PRECISÃO E ACURÁCIA PARA OS ESTIMADORES DA VARIÁVEL BIOMASSA DE ÁRVORES DE ACÁCIA NEGRA UTILIZANDO-SE MODELOS DE MÁQUINAS DE VETORES DE SUPORTE, PARA A BASE DE DADOS DE AJUSTE 670/2014.....	97
TABELA 14 – RESULTADOS DE VIÉS, PRECISÃO E ACURÁCIA PARA OS ESTIMADORES DA VARIÁVEL BIOMASSA DE ÁRVORES DE ACÁCIA NEGRA UTILIZANO-SE MODELOS DE MÁQUINAS DE VETORES DE SUPORTE APÓS UM REFINAMENTO, PARA A BASE DE DADOS DE AJUSTE 670/2014.....	97
TABELA 15 – RESULTADOS DE VIÉS, PRECISÃO E ACURÁCIA PARA OS ESTIMADORES DA VARIÁVEL BIOMASSA DE ÁRVORES DE ACÁCIA NEGRA UTILIZANDO MODELOS DE MÁQUINAS DE VETORES DE SUPORTES, PARA A BASE DE DADOS DE VALIDAÇÃO 98/2012.....	100
TABELA 16 – COMPARAÇÃO DE VIÉS, PRECISÃO E ACURÁCIA PARA OS MELHORES ESTIMADORES DE MÁQUINA DE VETORES DE SUPORTE PARA A BASE DE DADOS DE ESTIMAÇÃO E DE VALIDAÇÃO, EM RELAÇÃO AOS RESULTADOS OBTIDOS NA BASE DE DADOS DE UTILIZADA NA CRIAÇÃO DOS ESTIMADORES. .	101
TABELA 17 – COMPARAÇÃO DE VIÉS, PRECISÃO E ACURÁCIA PARA OS MELHORES ESTIMADORES DE BIOMASSA DE ÁRVORES DE ACÁCIA NEGRA MODELADA POR MEIO DE TÉCNICAS DE APRENDIZAGEM DE MÁQUINA, PARA A BASE DE DADOS DE AJUSTE 670/2014. ....	104

TABELA 18 – COMPARAÇÃO DE VIÉS, PRECISÃO E ACURÁCIA PARA OS MELHORES ESTIMADORES DE BIOMASSA DE ÁRVORES DE ACÁCIA NEGRA MODELADA POR MEIO DE TÉCNICAS DE APRENDIZAGEM DE MÁQUINA, PARA A BASE DE DADOS DE VALIDAÇÃO 98/2012.....	104
TABELA 19 – HIPERPARÂMETROS DOS ALGORITMOS DE APRENDIZAGEM DE MÁQUINA AJUSTADOS PARA A MODELAGEM DA VARIÁVEL BIOMASSA PARA A ESPÉCIE ACÁCIA-NEGRA E TAMBÉM PARA OUTRAS.....	107
TABELA 20 – RESULTADOS DE VIÉS, PRECISÃO E ACURÁCIA PARA OS ESTIMADORES DA VARIÁVEL BIOMASSA DE ÁRVORES DE ACÁCIA NEGRA UTILIZANDO MODELOS DE REGRESSÃO, USANDO A BASE DE DADOS DE AJUSTE 670/2014.....	109
TABELA 21 – RESULTADOS DE VIÉS, PRECISÃO E ACURÁCIA PARA OS ESTIMADORES DA VARIÁVEL BIOMASSA DE ÁRVORES DE ACÁCIA NEGRA UTILIZANDO MODELOS DE REGRESSÃO, PARA A BASE DE DADOS DE VALIDAÇÃO 98/2012.....	112

## LISTA DE ABREVIATURAS OU SIGLAS

AGEFLOR	- Associação Gaúcha de Empresas Florestais
ANOVA	- Análise de Variância
BNDES	- Banco Nacional de Desenvolvimento Econômico e Social do Brasil
CIBIO	- Congresso Internacional de Biomassa
INPE	- Instituto Nacional de Pesquisas Espaciais
MLP	- <i>Multilayer Perceptron</i>
REPL	- Read <i>Evaluate Print Loop</i>
RFF	- Resources for the Future
WEKA	- Waikato Environment for Knowledge Analysis
WNSUR	- <i>weighted-nonlinear seemingly unrelated regression</i>

## LISTA DE SÍMBOLOS

© - copyright

@ - arroba

$\Sigma$  - somatório de números

$\Pi$  - produtório de números

$i$  - idade

$d$  - diâmetro à altura do peito

$h$  - altura total

$b$  - biomassa total

$\bar{b}$  - média da biomassa total

$b_{est}$  - biomassa estimada

$\gamma$  - parâmetro gamma

$f_a$  - função matemática

$w$  - peso das variáveis

$x$  – variável

$u$  - união

$\partial$  - derivada parcial

$t_a$  - taxa de aprendizado

$B$  - número de variáveis

$d_{euclidiana}$  - distancia euclidiana

$\forall$  - quantificação universal

$\in$  - pertence

$\xi_i$  - variável folga

$C$  - constante

$n$  – número de árvores

$\#$  - caractere que denota a entrada da linguagem de programação *Julia* em seu modo de comentários

$\bar{e}$  - viés absoluto

$\bar{e}_{\%}$  - viés relativo

$S_e$  - precisão absoluta

$S_{e\%}$  - precisão relativa

$m_x$  - acurácia absoluta



$m_{x_{0\%}}$  - acurácia relativa

$e_i$  - erro sistemático para i-ésima árvore

IC - intervalo de confiança

## SUMÁRIO

<b>1 INTRODUÇÃO .....</b>	<b>24</b>
1.1 DEFINIÇÃO DO PROBLEMA.....	26
1.1.1 Modelagem da biomassa de árvores com aprendizagem de máquina em linguagem de programação <i>Julia</i> .....	26
1.2 HIPÓTESES .....	28
1.3 OBJETIVOS .....	29
1.3.1 Objetivo geral .....	29
1.3.2 Objetivos específicos.....	29
<b>2 REVISÃO DE LITERATURA .....</b>	<b>30</b>
2.1 BIOMASSA FLORESTAL.....	30
2.2 INTELIGÊNCIA ARTIFICIAL .....	31
2.2.1 Métodos de aprendizagem de máquina .....	32
2.2.1.1 Rede Neural Artificial .....	34
2.2.1.2 Floresta aleatória .....	37
2.2.1.3 <i>k</i> -vizinhos mais próximos.....	41
2.2.1.4 Máquina de vetores de suporte .....	43
2.2.1.5 Linguagem de programação <i>Julia</i> .....	46
<b>3 MATERIAL E MÉTODOS .....</b>	<b>49</b>
3.1 OBJETO DE ESTUDO .....	49
3.2 BASE DE DADOS DA ESPÉCIE ACÁCIA-NEGRA .....	50
3.2.1 Base de dados 670/2014 para ajuste dos algoritmos.....	51
3.2.2 Base de dados 98/2012 para validação do ajuste dos algoritmos.....	53
3.3 <i>DOWNLOAD</i> , INSTALAÇÃO DA LINGUAGEM DE PROGRAMAÇÃO <i>JULIA</i> E CARREGAMENTO DA BASE DE DADOS.....	54
3.4 ANÁLISE EXPLORATÓRIA DAS BASES DE DADOS .....	56
3.5 TÉCNICAS DE APRENDIZAGEM DE MÁQUINA USANDO A LINGUAGEM DE PROGRAMAÇÃO <i>JULIA</i> .....	57
3.5.1 Técnicas de aprendizagem de máquina.....	58
3.5.1.1 Redes neurais artificiais.....	59
3.5.1.2 <i>k</i> -vizinhos mais próximos.....	60

3.5.1.3 Florestas aleatórias .....	60
3.5.1.4 Máquinas de vetores de suporte .....	61
3.6 EQUAÇÕES DE REGRESSÃO DE BIOMASSA – TÉCNICA TESTEMUNHA....	61
3.7 CRITÉRIOS DE AVALIAÇÃO DA QUALIDADE PREDITIVA DA BIOMASSA POR TÉCNICAS DE APRENDIZAGEM DE MÁQUINA E MODELOS DE REGRESSÃO .	62
<b>4 RESULTADO E DISCUSSÃO .....</b>	<b>64</b>
4.1 ANÁLISE EXPLORATÓRIA DAS BASES DE DADOS .....	64
4.1.1 Estatística descritiva.....	64
4.2 ESTATÍSTICAS VIÉS, PRECISÃO E ACURÁCIA E GRÁFICO DE DISPERSÃO.....	77
4.3 TÉCNICAS DE APRENDIZAGEM DE MÁQUINA USANDO LINGUAGEM DE PROGRAMAÇÃO <i>JULIA</i> .....	80
4.3.1 Redes neurais artificiais .....	80
4.3.1.1 Predição da base de dados para validação 98/2012 .....	86
4.3.2 <i>k</i> -vizinhos mais próximos .....	87
4.3.2.1 Predição da base de dados para validação 98/2012 .....	91
4.3.3 Florestas aleatórias .....	92
4.3.3.1 Predição da base de dados para validação 98/2012 .....	95
4.3.4 Máquinas de vetores de suporte .....	95
4.3.4.1 Predição da base de dados para validação 98/2012 .....	100
4.4 COMPARAÇÃO DAS TÉCNICAS DE APRENDIZAGEM DE MÁQUINA APLICADOS NA MODELAGEM DE BIOMASSA.....	103
4.4.1 Avaliação do viés, precisão e acurácia.....	104
4.5 TÉCNICA TESTEMUNHA.....	108
4.5.1.1 Predição da base de dados 98/2012 para validação .....	111
4.6 MODELAGEM POR MEIO DE APRENDIZAGEM DE MÁQUINA.....	112
4.7 LINGUAGEM DE PROGRAMAÇÃO <i>JULIA</i> NA APRENDIZAGEM DE MÁQUINA.....	115
4.7.1 Aspecto de Usabilidade.....	115
4.7.2 Aspecto de desempenho computacional.....	116
4.7.3 Aspectos gerais da linguagem de programação <i>Julia</i> .....	117
4.7.4 Linguagem de programação <i>Julia</i> na área florestal.....	120
4.8 AVALIAÇÃO DAS HIPÓTESES FORMULADAS NA PESQUISA .....	120
<b>5 CONCLUSÕES .....</b>	<b>123</b>

5.1 RECOMENDAÇÕES PARA TRABALHOS FUTUROS .....	123
<b>REFERÊNCIAS.....</b>	<b>124</b>

## 1 INTRODUÇÃO

A biomassa florestal refere-se a toda massa vegetal existente na floresta ou apenas condizente à fração arbórea. A importância em quantificar a biomassa nas florestas é destacada por diversos pesquisadores e está associada aos cenários das mudanças climáticas e fixação de carbono (SANQUETTA et al., 2018a; DOLÁCIO et al., 2019; MARTINS et al., 2019), ciclagem de nutrientes (VAN BICH et al., 2018; DOS SANTOS et al., 2019; MOMOLLI; SCHUMACHER, 2019), fins energéticos e também para avaliação do crescimento da floresta (BEHLING et al., 2018; DE OLIVEIRA et al., 2018; MOMO TAKOUDJOU et al., 2018; LAU et al., 2019). Ainda, a avaliação da biomassa é importante para os planejamentos industriais que demandam essa matéria-prima da floresta, assim como para silvicultores que geram renda com a venda da mesma.

Quantificar a biomassa de uma floresta é um desafio para os profissionais da área florestal, dada as dificuldades em mensurar essa variável. É um processo laborioso e com custo elevado em relação à outras variáveis biométricas em uma floresta (BEHLING et al., 2019; DURANTE et al., 2019; OTA et al., 2019), como por exemplo, diâmetro à altura do peito, altura, diâmetro de copa e o volume do fuste, isso para cada árvore. Por causa disso, o uso de modelos de regressão para estimar a biomassa a partir dessas variáveis é a prática comum (DEBASTINI et al., 2018; OTERO et al., 2018; FIGUEIREDO; FIGUEIREDO, 2019; SHERRY et al., 2019).

A regressão é a técnica padrão de ajuste de modelos para as variáveis dendrométricas. Entretanto, devido à sua baixa flexibilidade e rigidez, especialistas na área buscam alternativas que possam se beneficiar de técnicas avançadas na computação (como a aprendizagem de máquina), bem como fazer uso das variáveis de diferentes naturezas em suas bases de dados, o que não é possível com a modelagem baseada em regressão (SANQUETTA et al., 2011, DA SILVA et al., 2013, SANQUETTA et al., 2016a). Ainda, em diversas situações, esses pesquisadores relatam a dificuldade de atender condicionantes da análise de regressão, o que originou mais um motivo para a busca de métodos não paramétricos, dado pelas técnicas de aprendizagem de máquina.

Nos últimos anos, pesquisas na área florestal fizeram uso de novas alternativas para modelar a biomassa, buscando obter estimativas mais acuradas. A exemplo, destacam-se os trabalhos realizados por Gleason e Im (2012), Hunter et al.

(2013), Kankare et al. (2013), Zhu e Liu (2015), Montañó et al. (2017), Behling et al. (2019) e Azizi (2019). Nota-se, por meio dessas pesquisas, o uso de diferentes técnicas de Aprendizagem de Máquina, uma subárea da inteligência artificial com a finalidade de realizar tarefas por meio da experiência em problemas anteriores (MITCHELL, 1997).

Na área florestal, as técnicas de Aprendizagem de Máquina mais comumente utilizadas são: redes neurais artificiais (DOS REIS MARTINS et al., 2016; CAMPOS et al., 2017; COSTA FILHO, 2019), k-vizinhos mais próximos (SANQUETTA et al., 2018b; SILVA et al., 2018; COSTA FILHO, 2019), florestas aleatórias (SILVA et al., 2017; SCHILOWSKI et al., 2018; COSTA FILHO, 2019) e máquinas de vetores de suporte (CORDEIRO et al., 2015; BINOTI et al., 2016; COSTA FILHO, 2019). As aplicações dessas técnicas, frequentemente, são realizadas por meio dos *softwares* WEKA, NEUROFOREST e linguagem de programação R.

Recentemente, uma nova linguagem de programação foi desenvolvida, denominada *Julia*, com intuito de apresentar facilidade de uso para os profissionais não familiarizados com a programação de computadores e que precisam de bom desempenho no processamento dos dados. Sendo assim, uma das contribuições desse trabalho é explorar e possibilitar aos pesquisadores da área florestal o uso da linguagem de programação *Julia*, como forma de disponibilizar ao pesquisador da área (geralmente com pouca experiência na área de programação) uma linguagem de programação que possa fazer uso de técnicas de aprendizagem de máquina na computação científica para análise de dados.

Mesmo a Ciência Florestal apresentando *softwares* e linguagens de programação consolidadas para realizar seus procedimentos, alternativas são necessárias. Sendo a linguagem *Julia* promissora, apresenta características que tendem a suprir as necessidades dos pesquisadores.

Contudo, apesar dos esforços para a elaboração de materiais didáticos para o ensinamento da linguagem de programação *Julia*, como encontrado em Sherrington (2015), Joshi (2016), Joshi e Lakhanpal (2017), Salceanu (2018a), Voulgaris (2016) e Balbaert (2016), não há materiais direcionados para a aplicação dessa linguagem na área da Ciência Florestal, o que diminui sua atratividade para os Engenheiros Florestais, o que corrobora para a criação do presente trabalho.



## 1.1 DEFINIÇÃO DO PROBLEMA

Como anteriormente evidenciado, este trabalho une três problemas:

- i) Modelar a biomassa de árvores;
- ii) Realizar as modelagens empregando técnicas de aprendizagem de máquina;
- iii) Efetuar os processamentos utilizando a linguagem de programação *Julia*.

A expectativa com este trabalho é despertar a curiosidade nos Engenheiros Florestais para com as possibilidades que a linguagem de programação *Julia* fornecesse. Em sua maioria os Engenheiros Florestais enfrentam dificuldades de lidar com programação, especialmente porque disciplinas voltadas para computação são escassas em suas grades curriculares. Ao analisar a estimativa da biomassa, espera-se que a aprendizagem de máquina seja mais generalista que a regressão apresentada na técnica testemunha.

### 1.1.1 Modelagem da biomassa de árvores com aprendizagem de máquina em linguagem de programação *Julia*

Como já mencionado, a biomassa florestal tem sido estudada com propósitos variados, dentre eles, ciclagem de nutrientes, fins energéticos, aspectos funcionais da floresta e avaliação de seu crescimento e por fim, no âmbito de fixação de carbono devido os problemas enfrentados com os cenários de mudança climática mundial.

Trabalhos como o de Golley et al. (1978), Haag (1985), Fonseca et al. (2018), Ferreira et al. (2019) e Phillips et al. (2019) indicam que o acúmulo de biomassa difere entre os ecossistemas. Essa variação pode ser decorrente de fatores ambientais e dos próprios fatores inerentes à planta (BARNES et al., 1998), bem como das condições edafoclimáticas de cada ambiente (HAAG, 1985). O acúmulo de biomassa é impactado por fatores que afetam a fotossíntese e a respiração (KOZLOWSKI; PALLARDY, 1996), o que demonstra a complexidade em se analisar informações de estoque de biomassa de uma floresta.

Assim, obter dados de biomassa de uma floresta se torna oneroso e moroso, ao se comparar com outras variáveis biométricas de uma floresta, o que denota a

prática de serem feitas estimativas de suas informações. Dessa forma, a modelagem da biomassa de árvores tornou-se usual, utilizando variáveis dendrométricas de mais fácil obtenção. Assim, a aplicação da modelagem viabiliza os inventários de biomassa, uma vez que reduz essencialmente os custos.

Por sua natureza, a regressão linear analisa apenas relações lineares entre variáveis dependentes e independentes. Isto é, pressupõe que existe uma relação direta entre eles. O que nem sempre é correto, visto que relações entre variáveis dependentes como biomassa e variáveis independentes como diâmetro e altura são não-lineares. A regressão linear analisa uma relação entre a média da variável dependente e as variáveis independentes.

Os modelos não-lineares são capazes de acomodar uma grande variedade de funções médias, embora cada modelo não-linear apresente menor flexibilidade do que os modelos lineares. Outra dificuldade desses modelos é a falta de soluções analíticas para estimar os parâmetros. Sendo assim, a escolha do modelo é crucial. A escolha de um modelo com parâmetros com significado biológico faz com que o processo de escolha de valores iniciais seja mais fácil, porque os valores de partida podem geralmente ser facilmente determinada por inspeção visual dos dados.

Contudo, alternativas são necessárias em virtude das limitações encontradas nos modelos de regressão, geralmente tendem a serem pouco flexíveis, implicando nas análises estatísticas e de comparação (BASKERVILLE, 1972; SANQUETTA et al., 2011). Outra limitação é a complexidade em ajustar modelos com muitas variáveis independentes, comumente encontrado nos problemas da área florestal (SANQUETTA et al., 2016a).

Nessa direção, profissionais da área florestal estão dedicando esforços com o objetivo de melhorar a acurácia dos estimadores de biomassa bem como encontrar novas alternativas para modelagem. Isso tem feito que esses profissionais se aprofundem em áreas importantes da computação, como a programação de computadores, bem como subáreas avançadas, como otimização e aprendizagem de máquina.

Tradicionalmente, os Engenheiros Florestais têm realizado suas análises em *softwares* não livres, como *Excel*, *SPSS* e *SAS*. Recentemente, a linguagem de programação *R* tem ganhado destaque na área florestal. A linguagem de programação *R* é de livre acesso e código aberto, com uma gama extensa de recursos voltados para as mais variadas tarefas de análise de dados, criação e

edição de gráficos, entre outras. Por esses e outros aspectos positivos, na maioria das universidades tem-se optado pelo uso da linguagem de programação *R*. Outras linguagens de programação, como *C* e *Java*, também têm sido utilizadas, principalmente por engenheiros mais familiarizados com computação.

Neste trabalho será usado uma linguagem de programação pouco explorada na Ciência Florestal. Optou-se por não realizar comparações entre os aspectos positivos e negativos das linguagens. Visto que a ideia da linguagem de programação *Julia* não é ser uma competidora das demais, aliás, tem como objetivo maximizar os pontos positivos de cada uma e unir esforços para aperfeiçoar a computação numérica. Essa característica é notória quando se observa a integração com as bibliotecas de outras linguagens.

Conforme descrito por Joshi e Lakhanpal (2017), a linguagem de programação *Julia* consegue unir performance com programação dinâmica, ser de fácil leitura e codificação como a *Python* e apresentar bom desempenho como o *C*. Assim, existe uma expectativa que a linguagem ofereça uma combinação de desempenho e produtividade e uma perspectiva de facilitar a computação científica para análise de dados. Daí sua importância de aplicá-la na área florestal, que muitas vezes lida com modelagem em bases com um grande conjunto de dados, ou mesmo, problemas de otimização complexos. A aprendizagem de máquina exige esforço computacional visto a grande variedade de parâmetros. Sendo assim, a linguagem *Julia*

## 1.2 HIPÓTESES

Dado os resultados da avaliação de acurácia em outros trabalhos de pesquisa referente a aplicação de técnicas de aprendizagem de máquina foi proposta a seguinte hipótese: *Técnicas de aprendizagem de máquina estimam a biomassa de maneira mais acurada que os modelos de regressão, uma vez que são mais flexíveis, podendo conciliar variáveis quantitativas e qualitativas e possuem maior capacidade de extrair e generalizar as informações das bases de dados quando comparados com os modelos alométricos tradicionais.*

Dado que *Julia* é uma linguagem com tipagem dinâmica de alto nível e com foco em programação científica numérica, foi proposta a seguinte hipótese: *a linguagem de programação Julia é apropriada para ser utilizada na modelagem de variáveis dendrométricas realizada neste trabalho, visto que possui diversas funções*

*matemáticas e estatísticas, alta precisão dos valores numéricos, facilidade para integrar com bibliotecas de outras linguagens, os códigos possuem mecanismos para facilitar o seu entendimento, a ideia do programador é expressa utilizando poucas linhas de código e é eficiente em termos de tempo de processamento dos dados.*

### 1.3 OBJETIVOS

#### 1.3.1 Objetivo geral

O objetivo geral deste estudo foi avaliar técnicas de aprendizagem de máquina na modelagem de biomassa da espécie acácia-negra com uso da linguagem de programação *Julia* e comparar os resultados com equações de regressão.

#### 1.3.2 Objetivos específicos

- i. Aplicar e avaliar técnicas de Aprendizagem de Máquina: rede neural artificial, máquina de vetor de suporte, floresta aleatória e k-vizinhos mais próximos em dados de acácia-negra para estimar a biomassa total;
- ii. Comparar o melhor resultado de Aprendizagem de Máquina com equações de regressão, considerada como a técnica testemunha; e
- iii. Efetuar as modelagens utilizando a linguagem de programação *Julia*.

## 2 REVISÃO DE LITERATURA

### 2.1 BIOMASSA FLORESTAL

O crescimento da demanda por produtos florestais exige dos profissionais da área florestal o aprimoramento de técnicas para quantificar a quantidade da matéria prima em florestas plantadas. Pode-se destacar que a biomassa do fuste constitui ainda a principal matéria prima das florestas plantadas.

O crescimento de uma determinada floresta é o resultado do saldo de biomassa acumulada pela árvore (MACHADO; FIGUEIREDO FILHO, 2003). Martinelli et al. (1994) descrevem biomassa como a quantidade expressa em massa do material vegetal proveniente de uma floresta. Já Sanquetta et al. (2002) afirmam que o termo biomassa florestal pode significar a biomassa total existente na floresta ou apenas a fração arbórea da mesma, podendo ser empregado o termo fitomassa florestal ou fitomassa arbórea.

Em termos práticos, a biomassa florestal ainda apresenta a denotação para massa verde ou massa seca, sendo que, massa verde é considerada o material fresco amostrado, apresentando porcentagem de água; já massa seca, a biomassa é obtida após secagem do material em estufa (CALDEIRA, 2003).

A partição da biomassa nos componentes da árvore varia conforme a espécie, fase de desenvolvimento, estado nutricional, condições edafoclimáticas, assim como a constituição química (LARCHER, 2006; RATUCHNE et al., 2016). O primeiro autor ainda descreve que a biomassa arbórea é composta, aproximadamente, por 50% de carbono, 43% de oxigênio, 6% de hidrogênio e 1% por nutrientes.

Um importante aspecto da biomassa florestal é a sua quantificação, com a finalidade de saber a ciclagem de nutrientes, crescimento e produção, capacidade energética e informações para estudos de estoque de carbono (SANQUETTA et al., 2014b; RATUCHNE et al., 2016; BEHLING et al., 2019). Tais estudos possuem importância para auxiliar na tomada de decisões no manejo dos recursos florestais (PÁSCOA et al., 2004; SANQUETTA et al., 2014b; SANQUETTA et al., 2015).

A quantificação da biomassa para fins energéticos tem sido uma atividade importante para os Engenheiros Florestais, porque a sua utilização como fonte energética tem sido incentivada nas últimas décadas em razão de aspectos ambientais, econômicos e energéticos. Referente ao ambiente, existe a necessidade

de se mitigar as emissões de gases do efeito estufa, reduzir o consumo de fontes não renováveis e proteger as florestas nativas (FAO, 2011; PAYN et al., 2015). Pelo aspecto econômico e energético, fatores como as fortes oscilações nos preços dos combustíveis fósseis, a busca por maior segurança energética e o desenvolvimento tecnológico da exploração de recursos renováveis têm incentivado a geração de energia a partir da biomassa florestal oriunda de florestas plantadas (VITAL et al., 2013; ZHANG et al., 2015).

De acordo com Vieira et al. (2008), a quantificação da biomassa pode ser feita por métodos de análise destrutiva, com o abate das árvores e coleta de amostras, e também a análise não destrutiva, com o uso de métodos estimativos, sendo a regressão o mais usual. A modelagem alométrica relacionam a variável dependente biomassa, com variáveis comumente medidas em campo nos trabalhos de inventário florestal (SANQUETTA et al., 2014c).

Como alternativa para o método convencional de estimar biomassa, já consolidado, os métodos de regressão por aprendizagem de máquina vêm produzindo resultados promissores (CUNHA et al., 2015; MIGUEL et al., 2015; DOLÁCIO et al., 2019).

O método indireto para quantificar a biomassa de árvores é predominante em empresas e instituições de pesquisas. Dessa forma, tal informação suporta o planejamento e a tomada de decisão (ASSIS et al., 2015), comprovando a importância do método para planejamentos florestais. Nesse sentido, é importante que seja dada continuidade nos estudos de métodos para modelar a biomassa, como é o caso da aplicação das técnicas de aprendizagem de máquina.

## 2.2 INTELIGÊNCIA ARTIFICIAL

A Inteligência Artificial é definida como uma área de estudo preocupada com a construção e representação de entidades inteligentes (RUSSEL; NORVIG, 2009). Esse termo começou a ser empregado na década de 50, com o encontro de pesquisadores de diversas áreas, como físicos, matemáticos, biólogos, engenheiros e cientistas da computação, no *Dartmouth College* nos Estados Unidos. Contudo, o início dos estudos de Inteligência Artificial se deu em 1943 e constituiu-se na simulação de uma rede de neurônios artificiais (MCCULLOCH; PITTS, 1943).



Entre as ciências básicas, tem sido usada em matemática, química e em física quântica. Carrasquilla e Melko (2017) utilizaram redes neurais artificiais para ensinar um computador a reconhecer diferentes fases da matéria e localizar as transições de fase associadas. A física da matéria condensada é o estudo do comportamento coletivo de conjuntos infinitamente complexos de elétrons, núcleos, momentos magnéticos ou átomos. Essa complexidade se reflete no tamanho do espaço de configuração que cresce exponencialmente com o número de partículas e a Inteligência Artificial é uma ferramenta que pode trabalhar com grandes quantidades de dados (CARRASQUILLA; MELKO, 2017). No mercado corporativo, aplicações da Inteligência Artificial são encontradas em: finanças, medicina, meteorologia, agricultura, engenharias, design gráfico, setor de *web* e o setor industrial.

Na área florestal, inovações em pesquisas com o uso de Inteligência Artificial podem ser observadas nos exemplos: Nunes e Görgens (2016) modelou a conicidade de árvores em complexos mosaicos de vegetação; Montañó et al. (2017) modelou a biomassa de florestas tropicais; Sakici e Günlü (2018) usou atributos de povoamentos a partir de dados de satélite; Sanquetta et al. (2018b) para estimar o volume do fuste e Nitoslowski et al. (2019) usou a Inteligência Artificial no setor de paisagismo urbano. Ressalta-se que várias outras modelagens na Ciência Florestal foram realizadas, inclusive anteriormente a esse período por outros pesquisadores.

### 2.2.1 Métodos de aprendizagem de máquina

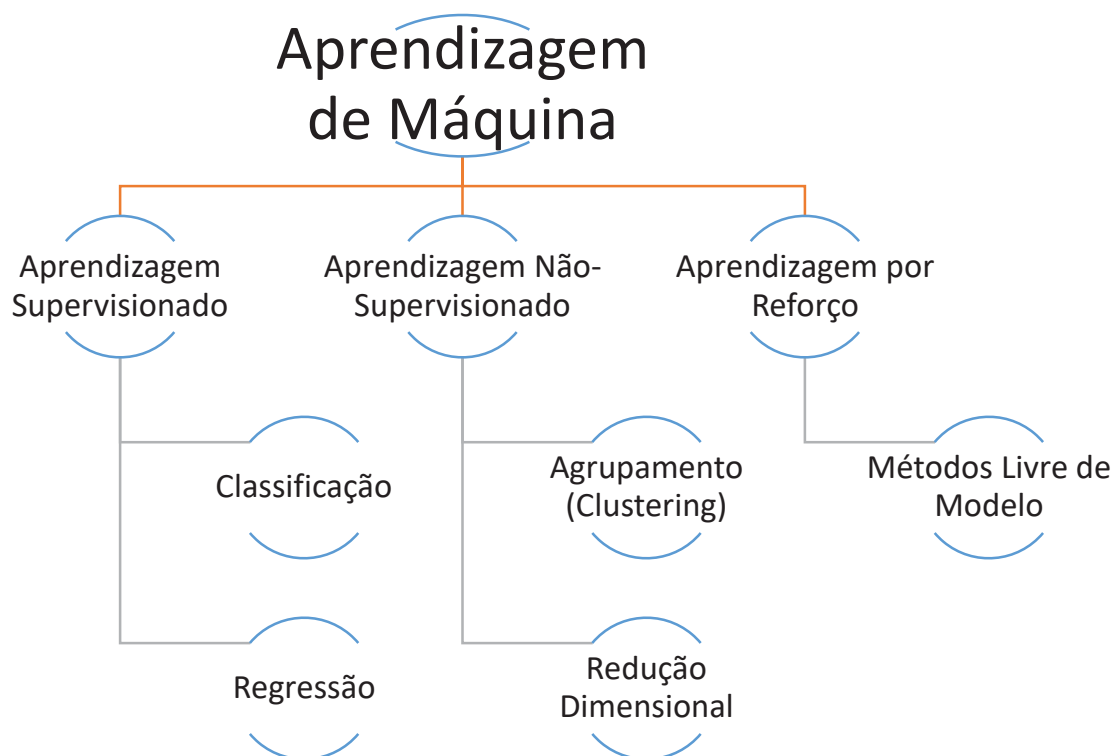
Russel e Norvig (2009) descrevem que a partir da década de 70 técnicas baseadas em Inteligência Artificial começaram a ser utilizadas de forma mais ampla na solução de problemas reais. Neste ponto, a Inteligência Artificial deixou de apresentar um viés puramente teórico para ganhar um aspecto prático. Boa parte dos problemas práticos resolvidos por técnicas de Inteligência Artificial eram, e ainda são, os que faziam uso de conhecimento de um determinado domínio para que se fossem feitas inferências, descobertas e conclusões a partir de informações fornecidas para o sistema.

Isso fez surgir uma subárea da Inteligência Artificial, denominada Aprendizagem de Máquina (MITCHELL, 1997) com viés ao estudo de técnicas computacionais sobre o aprendizado e a construção de sistemas capazes de agir de forma autônoma (MITCHELL, 1997; DUDA et al., 2000).

A aprendizagem de máquina pode ser vista como a inferência automática de alguns conceitos partindo de exemplos rotulados sobre o conceito a ser aprendido (MITCHELL, 1997). As técnicas de aprendizagem de máquina utilizam um princípio de inferência denominado de indução, em que as conclusões gerais emergem a partir de um conjunto de exemplos. Nesse âmbito, a aprendizagem supervisionada é um dos principais enfoques da indução, onde o sistema de aprendizagem deduz o conhecimento pela observação do seu ambiente (MALLOF et al., 2000).

Por outro lado, quando os dados não estão rotulados, a aprendizagem é denominada não supervisionada, e o objetivo é encontrar padrões usando alguma técnica de agrupamento ou associação. Na Figura 1 estão sumarizados alguns aspectos da aprendizagem de máquina (MITCHELL, 1997).

FIGURA 1 – APRENDIZAGEM SUPERVISIONADA E NÃO-SUPERVISIONADA



FONTE: O autor (2020).

A aprendizagem supervisionada consiste em uma operação para uma expectativa conhecida, algo que já se possui a resposta, dessa forma, a técnica de

aprendizagem buscará meios para chegar nas respostas. Mitchell (1997) define que os conjuntos de dados de entrada nesse contexto também são chamados de conjuntos de dados rotulados (*labeled data*). As técnicas classificadas nessa categoria se concentram no estabelecimento de uma relação entre os atributos de entrada e saída e usam essa relação especulativamente para gerar uma saída para novos pontos de dados de entrada.

Já na aprendizagem não-supervisionada um dos objetivos é encontrar a estrutura nos dados em relação ao mapeamento de construção entre os atributos de entrada e saída, sendo que os atributos de saídas não estão definidos. As técnicas definidas para essa função operam em um conjunto de dados não rotulados.

Na aprendizagem por reforço, usada principalmente na robótica e em problemas de controle, o aprendizado acontece por tentativa e erro (recompensas ou punições) ao atuar sobre um ambiente com propriedades de Markov. Desta maneira, não é necessária uma entidade externa que forneça exemplos ou um modelo a respeito da tarefa a ser executada: a única fonte de aprendizado é a própria experiência da máquina, cujo objetivo formal é adquirir uma política de ações que maximize seu desempenho geral (SUTTON; BARTO, 1998).

Dentre as técnicas para a geração de modelos de aprendizagem, destacam-se: máquina vetor de suporte, rede neural artificial, floresta aleatória e k-vizinhos mais próximos. Essas técnicas de aprendizagem são as mais aplicadas na resolução de problemas de aprendizagem supervisionada na área florestal, visto que suas características e flexibilidade possibilitam o uso e resultados satisfatórios na estimativa de variáveis dendrométricas<sup>1</sup> e por esses motivos serão usadas na presente pesquisa.

#### 2.2.1.1 Rede Neural Artificial

Uma rede neural artificial (*Artificial Neural Network*) é um modelo computacional biologicamente inspirado, constituído por elementos de processamento simples (nós ou neurônios artificiais) com a habilidade de aplicar determinada função matemática aos dados (função de ativação) produzindo uma resposta singular, sendo dispostos em camadas e ligados entre si, e essas conexões,

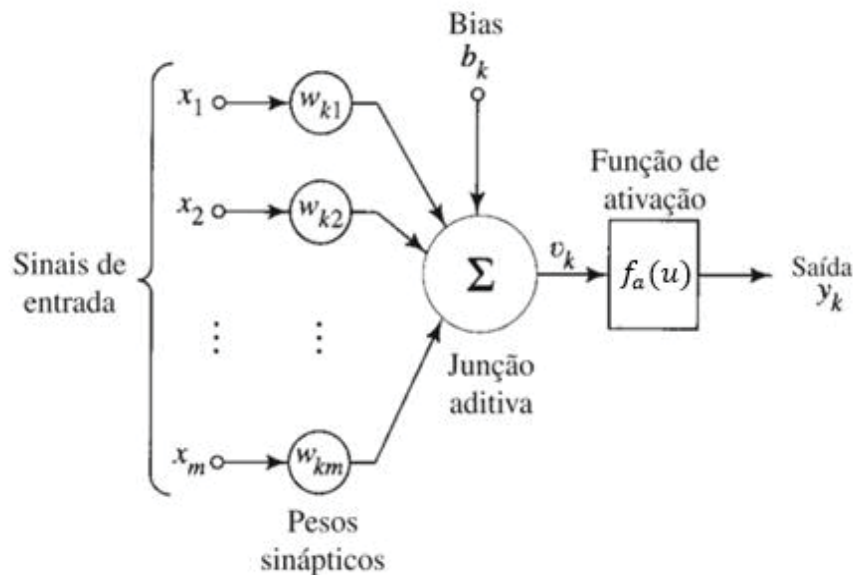
---

<sup>1</sup>Variável quantitativa oriunda da árvore (ex. diâmetro, altura, volume e biomassa).

geralmente, são associadas a coeficientes denominados de pesos (MCCULLOCH; PITTS, 1943; HOPFIELD, 1982; KOHONEN, 1982; HAYKIN; VEEN, 1999). Esses pesos são ajustados por um processo denominado treinamento ou aprendizagem, que fará a atualização dos pesos entre os neurônios (HAYKIN, 2001).

De modo geral, redes neurais artificiais necessitam de um período de treinamento, ajustes finos dos parâmetros e pode ser de difícil interpretação, muitas vezes não sendo possível identificar de forma clara a relação entre a entrada e a saída (HAYKIN; VEEN, 1999). A figura 2 ilustra os principais elementos que compõe um neurônio artificial.

FIGURA 2 – ESTRUTURA DE UM NEURÔNIO APRESENTANDO OS SEUS VALORES DE ENTRADA, SEUS PESOS, O SOMATÓRIO REALIZADO E A FUNÇÃO DE ATIVAÇÃO QUE RESULTA NA SUA SAÍDA.



FONTE: adaptado de Haykin (2001).

Haykin (2001) descreve que um neurônio recebe valores e retorna um resultado. Os valores de entrada são ponderados, combinados, somados e passam por uma função matemática  $f_a$ . Assim, sendo o vetor  $x = [x_1, x_2, \dots, x_m]^t$  a entrada de um neurônio e o vetor  $w = [w_1, w_2, \dots, w_m]^t$  representa os pesos aplicados a cada entrada, o resultado dos neurônio  $f'(x)$  é:

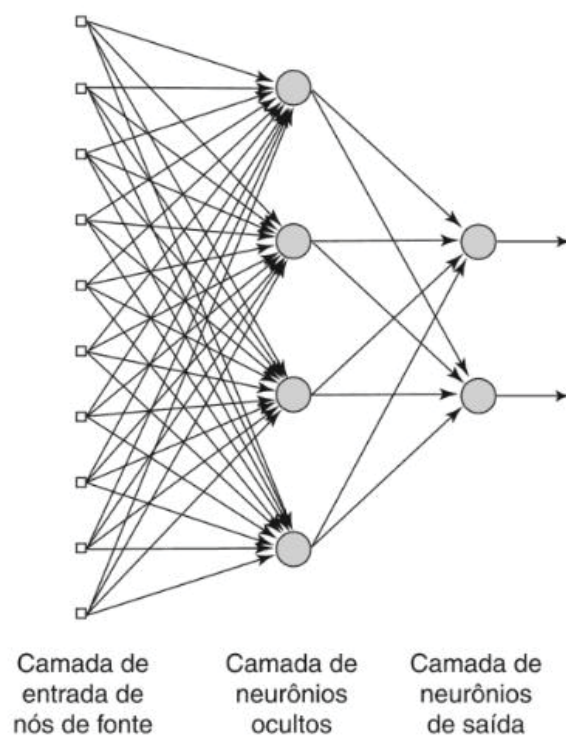
$$f_a(u) = \sum_{i=1}^m x_i \cdot w_i \quad (1)$$

$$f'(x) = f_a(u) \quad (2)$$

Haykin (2001) define que a função de ativação  $f_a$  pode ser de vários tipos, sendo as mais utilizadas a linear, limiar e sigmoideal. Funções lineares implicam no retorno de múltiplos de  $u$ . As funções limiares, que foram empregadas nos neurônios de McCulloch e Pitts (1934), definem quando o resultado é 0 ou 1 (ou -1), através de um valor limite. E a função sigmoideal representa uma aproximação contínua e diferenciável da função limiar.

Os neurônios podem estar dispostos em uma ou mais camadas. Um neurônio receberá como entrada as saídas dos neurônios da camada anterior e sua saída são colocadas na próxima camada. A camada do resultado é denominada de saída, já as demais camadas são ocultas (Figura 3) (HAYKIN, 2001).

FIGURA 3 – ARQUITETURA HIPOTÉTICA DE UMA REDE NEURAL ARTIFICIAL DE MÚLTIPLAS CAMADAS.



FONTE: Haykin (2001)

Cybenko (1989) afirma que para problemas não linearmente separáveis deve-se usar uma ou mais camadas ocultas. O autor ainda descreve que redes multicamadas, como a *Multilayer Perceptron* (MLP), usam funções não-lineares em suas camadas ocultas, como a função sigmoideal. Em problemas de classificação, cada neurônio de saída corresponde a uma classe, já em casos de regressão, a saída

não será discretizada e é retornado um valor decimal. Em geral, as MLP possuem cada neurônio de uma camada conectado a todos os neurônios da próxima camada.

Um dos algoritmos para aprendizagem de redes neurais artificiais MLP é o *backpropagation*, que ganhou popularidade na década de 80 (WERBOS, 1990). Esse algoritmo realiza o aprendizado pela correção de erros, em duas fases, a *forward* que se movimenta para frente e a *backward* que se movimenta para trás. Na fase *forward*, os neurônios calculam os valores com as ponderações específicas, e a função de ativação produz um valor de saída. O resultado da fase é checado com o resultado esperado e esta diferença é o erro que a rede apresenta. Já na fase *backward*, utiliza-se o erro cometido pela rede para ajustar os pesos dos neurônios (WERBOS, 1990; HAYKIN, 1998). O processo matemático é descobrir o quanto cada neurônio influencia no erro total da rede. Isto é, dado um peso  $w_i$  de alguma conexão na rede, calcula-se a derivada parcial do erro total da rede com respeito à  $w_i$  (3) e para cada peso  $w_i$  calcula-se o novo valor  $w_i^+$  (4).

$$w_i = \frac{\partial E_{total}}{\partial w_i} \quad (3)$$

$$w_i^+ = w_i - n \cdot \frac{\partial E_{total}}{\partial w_i} \quad (4)$$

Em que,  $n$  é a taxa de aprendizado.

As redes vêm sendo aplicadas na modelagem de relações envolvendo séries temporais complexas em várias áreas do conhecimento, tais como: evapotranspiração (DEO; SAHIM, 2015), previsão de safras agrícolas (EMAMGHLIZADEH *et al.*, 2015), geoestatística (DAI *et al.*, 2016), engenharia de pesca (YU *et al.*, 2006), radiação solar (LYRA *et al.*, 2016). Na área florestal, as aplicações são direcionadas a modelagem do volume do fuste de árvores (MARQUES DA SILVA *et al.*, 2009; BINOTI *et al.*, 2014; BINOTI *et al.*, 2015), altura (BINOTI *et al.*, 2013; VENDRUSCOLO *et al.*, 2015), afilamento (DOS REIS MARTINS *et al.*, 2016), volume (DOS SANTOS *et al.*, 2018; TAVARES JÚNIOR *et al.*, 2019) e biomassa (VAHEDI, 2016).

#### 2.2.1.2 Floresta aleatória

Floresta Aleatória é um modelo de aprendizagem baseado em árvores de decisão, que pode ser usado tanto para classificação e regressão (BREIMAN, 2001). O algoritmo para induzir uma floresta aleatória foi desenvolvido por Leo Breiman e Adele Cutler. O termo originou-se de florestas de decisão aleatória (*Random Decision Forest*) que foram propostas pela primeira vez por Tin Kam Ho, do laboratório Bell, em 1995.

Cada árvore utilizada na floresta aleatória é uma árvore de regressão que treina um número parametrizável de atributos, tomados de forma aleatória em cada treinamento, para sua construção. Sendo que essa árvore é conhecida como árvore aleatória (*Random Tree*) (BREIMAN, 2001).

Após o treinamento da floresta, as predições  $\hat{f}$  de valores não usados  $x'$  são feitas por meio da equação 5, que para regressão é a média dos valores obtidos em cada árvore da floresta.

$$\hat{f} = \frac{1}{N} \sum_{n=1}^N \hat{f}_b(x') \quad (5)$$

Em que  $N$  é o número de árvores.

Gollapudi (2016) define que para uma floresta aleatória com  $n$  árvores, o treinamento dos classificadores (ou regressores) da árvore de decisão é feito da seguinte forma:

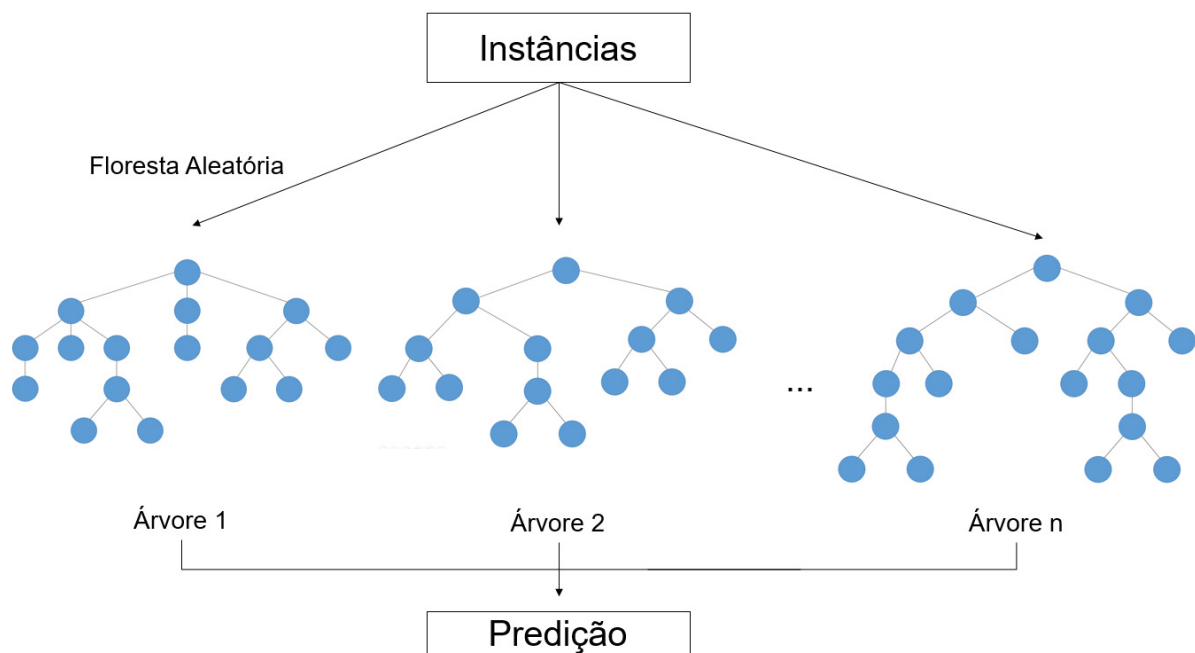
- Semelhante à técnica de ensacamento (*Standard Bagging Technique*), de modo que uma amostra de  $n$  compartimentos é definida com uma substituição aleatória para criar um subconjunto (*subset*) dos dados que representa cerca de 62 a 66% do conjunto abrangente.
- Para cada nó tem-se a necessidade:
  - De selecionar as variáveis preditoras  $m$  de modo que a variável identificada dê a melhor divisão, sendo essa uma divisão binária.
  - Para o próximo nó, deve-se escolher outra variável dentre as  $m$  variáveis para que o processo se repita no próximo nó.
- O valor de  $m$  pode variar:
  - Para seleção aleatória de divisores  $m$  assume o valor de 1.

- No método de ensacamento de Breiman (*Breiman's Bagging*)  $m$  assume o valor do número total de variáveis preditivas.
- Para florestas aleatórias,  $m$  é menor que o número de variáveis preditoras e pode levar três valores:  $1/2 \sqrt{m}$ ,  $\sqrt{m}$  e  $2\sqrt{m}$ .
- Dessa forma, para cada nova entrada na floresta aleatória para previsão, o novo valor será percorrido por todas as árvores e uma média, média ponderada ou maioria de votos é usada para obter o valor previsto.

Na Figura 4 está ilustrado uma representação de uma floresta aleatória, onde os valores de entrada são apresentados em todas as árvores da floresta. Os valores adquiridos são então consolidados pela da média, que é o valor final predito pela floresta. O intuito é aumentar a variância do conjunto e reduzir o viés dos estimadores, levando a um melhor modelo do que se fossem usadas árvores isoladas. Isso significa que enquanto previsões usando-se somente uma árvore são sensíveis a ruído no conjunto de treinamento, a média de várias árvores não é, por causa da não correlação entre elas.

FIGURA 4 – ESQUEMA DE FUNCIONAMENTO DE UMA FLORESTA ALEATÓRIA. A ENTRADA É APRESENTADA A TODAS AS ÁRVORES DA FLORESTA E O RESULTADO DA PREDIÇÃO É A MÉDIA DOS RESULTADOS DE CADA ÁRVORE.





FONTE: Adaptado de Breiman (2001).

Além de se conseguir amostras do conjunto de treinamento, a técnica de floresta aleatória também obtém amostras do conjunto de características, para elaboração da árvore. Este processo é chamado de *bagging* de características (*feature bagging*). A finalidade é reduzir a correlação entre as árvores geradas e para árvores de regressão de características, Hastie et al. (2003) recomenda 3 como número de características em cada árvore.

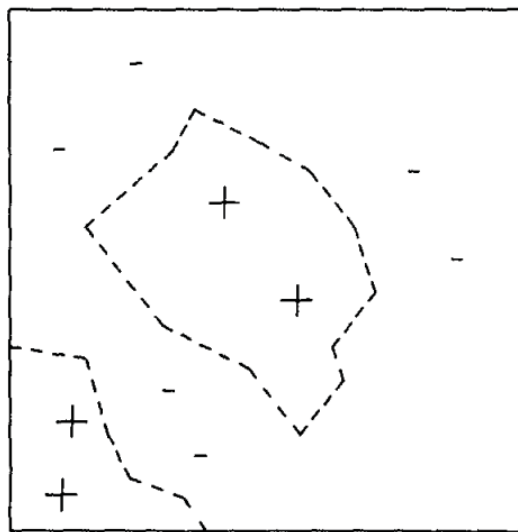
Segundo Bhuschjäger et al. (2018), a aplicação de florestas aleatórias tornou-se importante em várias áreas, como: classificação em tempo real de objetos celestiais em astrofísica, detecção de pedestres em tempo real, análise 3D de face em tempo real, classificação de sinais sonoros em tempo real, nano sensores, etc. Além de tais usos, os autores descrevem a aplicação do método para a avaliação em tempo real pelo esquadramento de árvores. Kilham et al. (2019) descrevem o uso de florestas aleatórias para gerar previsões de colheita a partir de inventários florestais.

Na área florestal são aplicadas em problemas de sensoriamento remoto (PEERBHAY et al., 2015; DEKLERCK et al., 2017; ORTIZ-REYES et al., 2019), ecologia (FOX et al., 2017; HART et al., 2017; YOUNG et al., 2018; BARNARD et al., 2019), incêndios florestais (GUO et al., 2016; JAAFARI; POURGHASEMI, 2019), modelagem das variáveis: biomassa (GHOSH; BEHERA, 2018; XING et al., 2019) e volume (DE ABREU et al., 2018; IQBAL et al., 2018).

### 2.2.1.3 $k$ -vizinhos mais próximos

O algoritmo  $k$ -vizinhos mais próximos (*k-Nearest Neighbors*) foi proposto por Cover e Hart (1966), demonstrando que era eficiente para resolver problemas de classificação de padrões. Outro motivo para o emprego desse algoritmo seria sua simplicidade e fácil implementação, contudo há presença de pontos fracos em sua estrutura, como: i) necessidade de grande capacidade de armazenamento; ii) o algoritmo requer muito tempo computacional; iii) possui sensibilidade a ruídos e falsos padrões (AHA et al., 1991). Esses autores descrevem o método como sendo um dos mais simples e intuitivos na área de aprendizagem de máquina, visto que o mesmo é um método não paramétrico onde uma nova observação recebe valor ou a classe mais próximo presente no conjunto de exemplos de treinamento.

FIGURA 5 – ESQUEMA DE FUNCIONAMENTO DE UM ALGORITMO DE  $k$ -VIZINHOS MAIS PRÓXIMOS



FONTE: Aha et al. (1991).

O algoritmo de vizinhos mais próximos tem como principal característica assegurar que o valor interpolado seja um valor original, dessa forma, ele não gera novos valores. O produto final desse interpolado é caracterizado por um efeito de degrau (HASTIE; TIBSHIRANI, 1996). A regra de decisão, desse algoritmo, fornece um procedimento não paramétrico simples, atribuindo um rótulo de classe ao padrão de entrada baseado nos rótulos apresentados nas proximidades, ou seja os vizinhos do vetor.

O método pode ser descrito como  $L = \{(y_i, x_i)\}$ , sendo o conjunto de dados de treinamento, onde  $y_i$  denota a variável dependente, e  $x_i$  as variáveis preditoras. A determinação dos vizinhos mais próximos é baseada em uma função de distância  $d(x_i, x_j)$ . O valor de nova observação  $(y, x)$  será o mesmo de seu vizinho mais próximo  $((y_{(1)}, x_{(1)}))$ , dessa forma, o vizinho mais próximo é aquele que possui a menor distância calculada da observação, dentre todas as observações empregadas no treinamento do algoritmo (HECHENBICHLER; SCHLIEP, 2004).

A distância é calculada considerando-se as métricas definidas, sendo alguns exemplos comumente usados: distância euclidiana, euclidiana quadrática, Chevychev ou a distância de Manhattan (DUDANI, 1976). A métrica de distância é um hiperparâmetro do algoritmo e deve ser definido para realização do treinamento. De modo a exemplificar, a seguir tem-se a formulação para cálculo da distância euclidiana entre duas observações.

$$d_{euclidiana}(x_i, x_j) = \sqrt{\sum_{s=1}^p (x_{is} - x_{js})^2} \quad (6)$$

Uma variação comum para esse método é o  $k$ -vizinhos mais próximos (*k-nearest neighbor*). Dessa forma, esse método não busca somente o exemplo mais similar para determinação do rótulo, mas também os  $k$  exemplos mais similares (HECHENBICHLER; SCHLIEP, 2004). Vale ressaltar que o parâmetro  $k$  deve ser determinado pelo próprio usuário.

Quando o hiperparâmetro  $k$  é igual a 1, o algoritmo corresponde ao método do vizinho mais próximo, e quando  $k$  é maior que 1, o rótulo do novo exemplo é determinado a partir da média entre os  $k$ -vizinhos mais próximos para os problemas de regressão (AHA et al., 1991).

Na área florestal pode-se observar o uso de tais algoritmos nas áreas: ecologia (THOMAS et al., 2016; RUIVO et al., 2017; BELLISARIO et al., 2019), incêndios florestais (KERN et al., 2017; KAUR; SOOD, 2019; WATSON et al., 2019), sensoriamento remoto (CHIRICI et al., 2016; ALIMJAN et al., 2017; FINLEY et al., 2019), serraria (ZAMAR et al., 2017; OGUNTOKE et al., 2019 MORIN et al., 2019) e

modelagem de variáveis dendrométricas: biomassa (MCROBERTS et al., 2015; SANQUETTA et al., 2018b), volume (RÄTY et al., 2019; SOUZA et al., 2019).

#### 2.2.1.4 Máquina de vetores de suporte

Máquinas de vetores suporte (*Support Vector Machines*) baseiam-se em técnicas embasadas na Teoria de Aprendizado Estatístico (BURGER, 1998). Os vetores de suporte são técnicas vastamente utilizadas em problemas de classificação e regressão. Seu princípio fundamental é fazer correções por meio de mapeamento não-linear dos conjuntos de dados de entradas para um espaço de alta dimensão, usando um conjunto de funções Kernel (CHEN et al., 2011; VAPNIK, 2013; GUENTHER; SCHONLAU, 2016; ZHANG-JAMES et al., 2019).

Máquinas de vetores de suporte são robustas diante de dados de grande dimensão e superior a outras técnicas. Em virtude de haver a convexidade o problema da otimização formulada em seu treinamento, que implica na existência de um único mínimo global (SHRESTHA; SHUKLA, 2015).

As técnicas de vetores de suporte destacam-se por sua boa capacidade de generalização. Álvares Meza et al. (2012) descreve que em regressão, duas abordagens foram criadas: regressão de vetores de suporte (*support vector regression*) e máquinas de vetores de suporte de mínimos quadrados (*least-squares support vector machines*). Sendo que para os estudos desse trabalho, optou-se pelo primeiro método.

Burger(1998) define que de modo a compreender os princípios fundamentais das máquinas de vetores de suporte, pode-se admitir um conjunto de dados de entrada  $(x_i, y_i)$ , em que  $x_i \in R$  é um vetor que pertence a duas classes  $y_i \in \{-1; +1\}$  (Figura 7A). A proposta do vetor de suporte é dividir as classes por meio de um hiperplano, expresso por:

$$f(x) = w.x + b \quad (7)$$

Em que:  $w$  é o conjunto de parâmetros e  $b$  o desvio aleatório.

Pode-se considerar a existência de inúmeras soluções possíveis (mínimos locais) que separem o conjunto de dados em dois (Figura 6), dessa forma, a técnica de máquinas de vetores de suporte emprega o princípio da minimização do risco

estrutural, visando obter a obtenção do hiperplano de máxima distância das amostras de classificação, conhecido como hiperplano de separação ótima. Por fim, são denominados vetores de suporte as amostras que delimitam as margens de separação entre as classe, as quais assumem valor  $w.x + b = +1$  quando pertencentes a classe  $+1$ , e  $w.x + b = -1$  quando pertencentes à classe  $-1$  (BURGER, 1998).

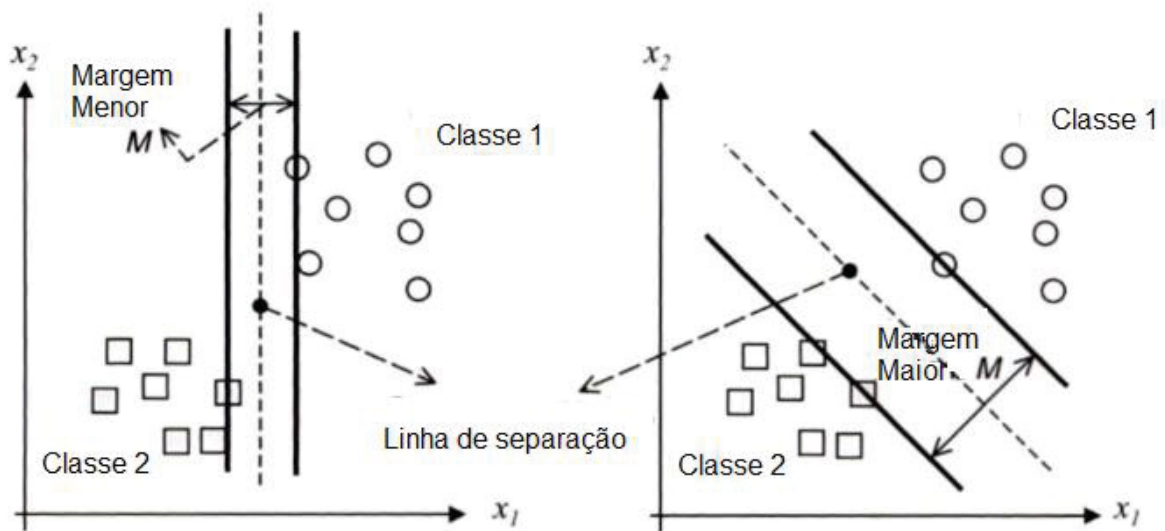
Para otimização da função (7) tem-se a descrição do seguinte problema de otimização:

$$\text{minimize } \frac{1}{2} \|w\|^2 \quad (8)$$

$$\text{sujeito a } \begin{cases} w.x + b \geq +1 \forall y \in \{+1\} \\ w.x + b \leq -1 \forall y \in \{-1\} \end{cases} \quad (9)$$

Em que  $\|w\|^2$  é o produto escalar do vetor  $w$  com ele mesmo, ou seja  $(w.w)$ .

FIGURA 6 – ILUSTRAÇÃO DE UM PROBLEMA DE CLASSIFICAÇÃO BINÁRIA LINEARMENTE SEPARÁVEL, COM SOLUÇÕES POSSÍVEIS PARA O PROBLEMA DE SEPARAÇÃO DOS CONJUNTOS ATRAVÉS DE UM HIPERPLANO



FONTE: Wang(2005).

No caso apresentado por Wang (2005) as amostras são linearmente separáveis, contudo, na grande totalidade das vezes, duas classes não são separáveis por um hiperplano, o que acarretou no desenvolvimento do conceito de

variável de folga  $\xi_i$ , por Cortes e Vapnik (1995), o que permitiu a existência de erros de classificação.

Sendo assim, Burger (1998) descreve que o erro é proporcional à distância da amostra à margem definida pelos vetores de suporte da classe à qual pertencem. O autor ainda menciona que para controlar os erros de classificação acarretados pela flexibilização da largura entre margens adicionou-se a constante  $C_1$ , sendo, que quando a constante apresenta valor baixo admite que mais amostras possam ser classificadas erroneamente, enquanto um valor alto torna o problema menos tolerante a erros. O valor do parâmetro  $C$  está relacionado à complexidade de uma técnica de vetor de suporte, regulando o subajuste e o sobreajuste (HUANG et al., 2006).

A constante  $C$  é acrescentada ao problema de otimização juntamente à variável  $\xi_i$ , para controlar a tolerância à erros, modificando a função objetivo que toma as seguintes expressões:

$$\text{minimize } \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \xi_i \quad (10)$$

$$\text{sujeito a } \begin{cases} w \cdot x + b \geq +1 - \xi_i \\ w \cdot x + b \leq -1 - \xi_i \\ \xi_i > 0 \end{cases} \quad (11)$$

Na literatura há diversos exemplos de aplicações de vetores de suporte em áreas distintas de investigação, tais como: estudo de mercado (CASTAÑEDA, 2019), predição da expansão urbana (KARIMI et al., 2019), análise de cenas (SHARIF et al., 2019), modelagem da evapotranspiração de referência e de cultura (KISI; ÇIMEN, 2009; TABARI et al., 2012; SHRESTHA; SHUKLA, 2015), presciência dos graus de concentração de ozônio troposférico (FENG et al., 2011; LUNA et al., 2014), qualidade do papel utilizando eucalipto (NIETO et al., 2012), investigação da propagação de algas em água doce (LOU et al., 2017), ponderação do tempo de sono de novilhos em sistemas de produção de gado (HOKKANEN et al., 2011), previsões diárias e mensais de vazões em reservatórios nos Estados Unidos (LIU et al., 2014), modelagem da distribuição potencial de espécies por meio de condições bióticas e abióticas onde o conhecimento das espécies pode ocorrer (LORENA et al., 2011), teste com sucesso de modelos gerados com dados de entrada de temperatura do ar para modelar de irradiação solar global (HG) (CHEN et al., 2011).

Na área florestal as técnicas de máquinas de vetores de suporte vêm sendo empregados nas áreas de: ecologia (BOUREL; SEGURA, 2018; LIU et al., 2018; BELISSARIO; PIJANOWSKI, 2019), incêndios florestais (KERN et al., 2017; CAI et al., 2019; JAAFARI; POURGHASEMI, 2019), sensoriamento remoto (CAVALLARO et al., 2015; MAULIK; CHAKRABORTY, 2017; SILVASANKAR et al., 2019) e modelagem de variáveis dendrométricas: biomassa (SHARIFI et al., 2017; MALEK et al., 2019), altura total (COSTA FILHO et al., 2019) e volume (DE ABREU et al., 2018; LOPES et al., 2018; WANG et al., 2019).

### 2.2.1.5 Linguagem de programação *Julia*

No ano de 2009, três programadores (Stefan Karpinski, Jeff Bezanson e Viral Shah) participantes do grupo de computação aplicada começaram a trabalhar em um projeto que futuramente se tornaria a linguagem de programação *Julia*, com a supervisão do Professor Alan Edelman (JOSHI, 2016).

Para Bezanson *et al.* (2012) as características mais importantes da linguagem de programação *Julia* concerne:

- i) A biblioteca padrão é escrita utilizando a própria linguagem de programação *Julia*, incluindo operadores primitivos como operações aritméticas de inteiros.
- ii) Uma grande variedade de tipos para construir e descrever objetos, que pode também, opcionalmente, ser utilizado para fazer declarações de tipos.
- iii) A habilidade de definir o comportamento de funções com base na combinação de vários tipos de argumentos via despacho múltiplo.
- iv) Geração automática de código eficiente e especializado para diferentes tipos de argumentos.
- v) Bom desempenho, aproximando-se de linguagens estatísticas e compiladas como a linguagem de programação C.

Os desenvolvedores da linguagem de programação destacam:

*We want the speed of with the dynamism of Ruby. We want something as usable for general programming as Python, as easy for statistics as R, as natural for string processing as Perl, as powerful for linear algebra as Matlab...*  
(BEZANZON et al., 2012).



Porém, por ser uma linguagem de programação nova, ao se comparar com as linguagens já consolidadas com mais de 30 anos de aprimoramento, sua literatura ainda é pouco desenvolvida. Vários pesquisadores estão dedicando esforços para desenvolver materiais e incluir a linguagem de programação *Julia* nas universidades, principalmente em países como Estados Unidos, Inglaterra, Alemanha, China, Japão, Áustria, entre outros. A versão atual no *GitHub* conta com 45.885 commits, 908 contributores, 25.066 stars e 9.318 pacotes registrados.

No Brasil, a utilização da linguagem de programação *Julia* em universidades já começou, destacando-se o uso por pesquisadores da Universidade de São Paulo, Universidade Federal de Viçosa, Universidade Federal do Paraná, Universidade Federal de Alagoas e Universidade Federal de Santa Maria. Ainda, a aplicação pelo Banco Nacional de Desenvolvimento Econômico e Social do Brasil (BNDES) e Instituto Nacional de Pesquisas Espaciais (INPE), entre outros.

Os principais livros publicados nos últimos 4 anos foram escritos por: Balbaert (2015), Sengupta (2016), Vougaris (2016), Joshi e Lakhanpal (2017), Nagar (2017), Balbaert (2018), Kaminski e Szufel (2018), Salceanu (2018b), Sengupta (2018), Balbaert e Salceanu (2019), Kwon (2019), Lauwens e Downey (2019) e Sherrington (2019).

Balbaert (2015), Joshi e Lakhanpal (2017), Lauwens e Downey (2019) e Nagar (2017) publicaram livros focando no entendimento do sistema da linguagem de programação *Julia* e na criação de programas simples. Apresentaram como criar seus próprios sistemas dentro da linguagem e como definir funções e entender a metaprogramação. Abordam também a criação de visualizações gráficas e dados e elaboração de programas capazes de trabalhar em rede de computação paralelas.

Vougaris (2016), Balbaert (2018), Kaminski e Szufel (2018), Sengupta (2018), Balbaert e Salceanu (2019), Sherrington (2019) redigiram livros com enfoque mais estatístico e aplicações de *big data*. Apresentam, também, como executar programas externos e *packages* dentro da linguagem de programação *Julia*.

Sengupta (2016) apresentou *Julia High performance*, o qual visa ensinar a codificar programas de alta confiabilidade e alto desempenho, trazendo soluções para as complexidades enfrentadas pelos desenvolvedores enquanto desenvolve código elegantes. Isso conduz o leitor a entender as características de desempenho dos seus programas em linguagem de programação *Julia* e permitir que o mesmo utilize a promessa de níveis próximos da linguagem de programação C. O leitor aprende a



analisar e medir o desempenho do código da linguagem de programação *Julia*, entendendo como evitar gargalos e melhor projetar um programa.

Além de materiais acadêmicos e científicos, encontros e conferências a respeito dessa linguagem de programação tem denotado atenção. Dentre eles, destaca-se a *JuliaConque* é a maior conferência a respeito de linguagem de programação *Julia*, reunindo palestrantes, desenvolvedores e entusiastas de todo o mundo. A conferência ocorre anualmente desde 2014, em um período de 5 dias, com palestras e workshops. No ano de 2019 a conferência ocorreu em Baltimore nos Estados Unidos e em 2020 *JuliaCon* ocorrerá em Lisboa, Portugal.

Na área florestal, essa linguagem ainda está pouco difundida, contudo, já pode-se notar mudanças nesse viés. No Congresso Internacional de Biomassa (CIBIO), que ocorreu em Curitiba, no ano de 2019, foi denotado um espaço para uma palestra sobre *Julia*, enfatizando o aumento do interesse e da visibilidade da linguagem na área e a relevância de estudos aplicados. Ainda, ressalta um projeto de pesquisa que está sendo desenvolvido no curso de Engenharia Florestal da Universidade Federal do Paraná intitulado: Linguagem *Julia* para Engenheiros Florestais. O projeto tem objetivo de realizar aplicações da linguagem em 3 campos importantes da Ciência Florestal: Experimentação Florestal, Inventário Florestal e modelagem por meio de aprendizagem de máquina. Ainda, projetos desenvolvidos na Universidade Federal de Santa Maria e Universidade Federal de Viçosa voltados ao processamento de imagens *LiDAR* e também para análise de dados de melhoramento genético.

Portanto, a contribuição desse trabalho é unir áreas (Ciência Florestal – biomassa, Estatística – modelagem da biomassa e Computação – linguagem de programação) e assim desenvolver conhecimento a respeito da área florestal, criando estimadores por aprendizagem de máquina para biomassa de árvores de acácia-negra, e para isso utilizar uma linguagem de programação alternativa e promissora intitulada *Julia* que apresenta pouca base para programação na ciência florestal.

### 3 MATERIAL E MÉTODOS

#### 3.1 OBJETO DE ESTUDO

A espécie acácia-negra (*Acacia mearnsii* De Wild.) (Figura 7) é o objeto de estudo da presente pesquisa, sendo originária da Austrália e pertencente à família *Fabaceae*. Seu nome popular faz menção à coloração de sua casca e folhagem (MARTINS et al., 2019).

No estado do Rio Grande do Sul - Brasil, são cultivados três principais gêneros florestais *Acacia*, *Eucalyptus* e *Pinus*. O estado apresentou no ano de 2016, uma área total plantada de 780,9 mil hectares, sendo que deste montante, 89,6 mil foram de plantas do gênero *Acacia*, demonstrando que o estado é detentor de quase 100% do plantio de acácia no país (AGEFLOR, 2017). A Associação Gaúcha de Empresas Florestais (AGEFLOR) ainda descreve que a acácia-negra é a principal espécie do gênero presente no estado do Rio Grande do Sul, sendo que suas florestas plantadas são uma fonte de matéria-prima para a indústria (taninos, celulose e carvão), além da importância social ao contribuir para a renda direta e indiretamente de mais de 35 mil famílias.

FIGURA 7 – POVOAMENTO DE ACÁCIA-NEGRA COM IDADES DE 7 E 5 ANOS, RESPECTIVAMENTE, NO ESTADO DO RIO GRANDE DO SUL, BRASIL.



FONTE: O autor (2020).

Nos povoamentos em que foi realizada a presente pesquisa, os plantios foram cultivados manualmente com mudas produzidas a partir de sementes oriundas de pomares da empresa TANAGRO S.A. Para cada muda, foram utilizados 50 gramas de adubo NPK (5-30-15). O espaçamento utilizado foi de 3 m entre linhas e 1,75 m

entre plantas na linha. O manejo empregado foi igual para todos os povoamentos, com a realização de roçada mecanizada com trator e roçadeira agrícola para controle das plantas daninhas durante os dois primeiros anos após o plantio.

### 3.2 BASE DE DADOS DA ESPÉCIE ACÁCIA-NEGRA

Foram utilizadas duas bases de dados da espécie acácia-negra denominadas de:

- i) Base de dados 670/2014 para treinamento das técnicas de aprendizagem de máquina; e
- ii) Base de dados 96/2012 para validação do ajuste das técnicas de aprendizagem.

Essas duas bases de dados foram coletadas nos mesmos locais de pesquisa, em uma mesma variação de idade e com o mesmo método de mensuração das variáveis, porém em diferentes povoamentos e época. A base de dados para treinamento foi composta por 670 árvores e estabelecida no ano de 2014. Ressalta-se que são poucas as pesquisas que contemplam esse tamanho de amostra para a variável biomassa de uma mesma espécie. A base de dados para validação foi composta por 96 árvores e estabelecida no ano de 2012, correspondendo a um tamanho de mais de 10% da primeira amostra, o que se julga apropriado para o propósito.

Para a modelagem por meio de técnicas e algoritmos de aprendizagem de máquina, foram utilizadas as variáveis independentes diâmetro à altura do peito ( $d$ )<sup>2</sup>, altura total ( $h$ )<sup>3</sup> e idade ( $i$ ) e, como variável dependente, a biomassa total acima do solo ( $b$ )<sup>4</sup>. Os resultados obtidos foram comparados com uma técnica testemunha, dada pelas equações de regressão propostas por Behling et al. (2019) para a base de dados 670/2014, que utilizaram como variáveis independentes  $d$  e  $h$ , sendo equações ajustadas em estratos de grupos de idade dos povoamentos.

---

<sup>2</sup>É convencionalizado como a medida do diâmetro obtida a 1,30m acima do solo.

<sup>3</sup>Altura máxima, medindo o ponto mais baixo, que toca o solo, até a ponta da copa.

<sup>4</sup>Mensuração da biomassa que exclui os componentes abaixo do solo.

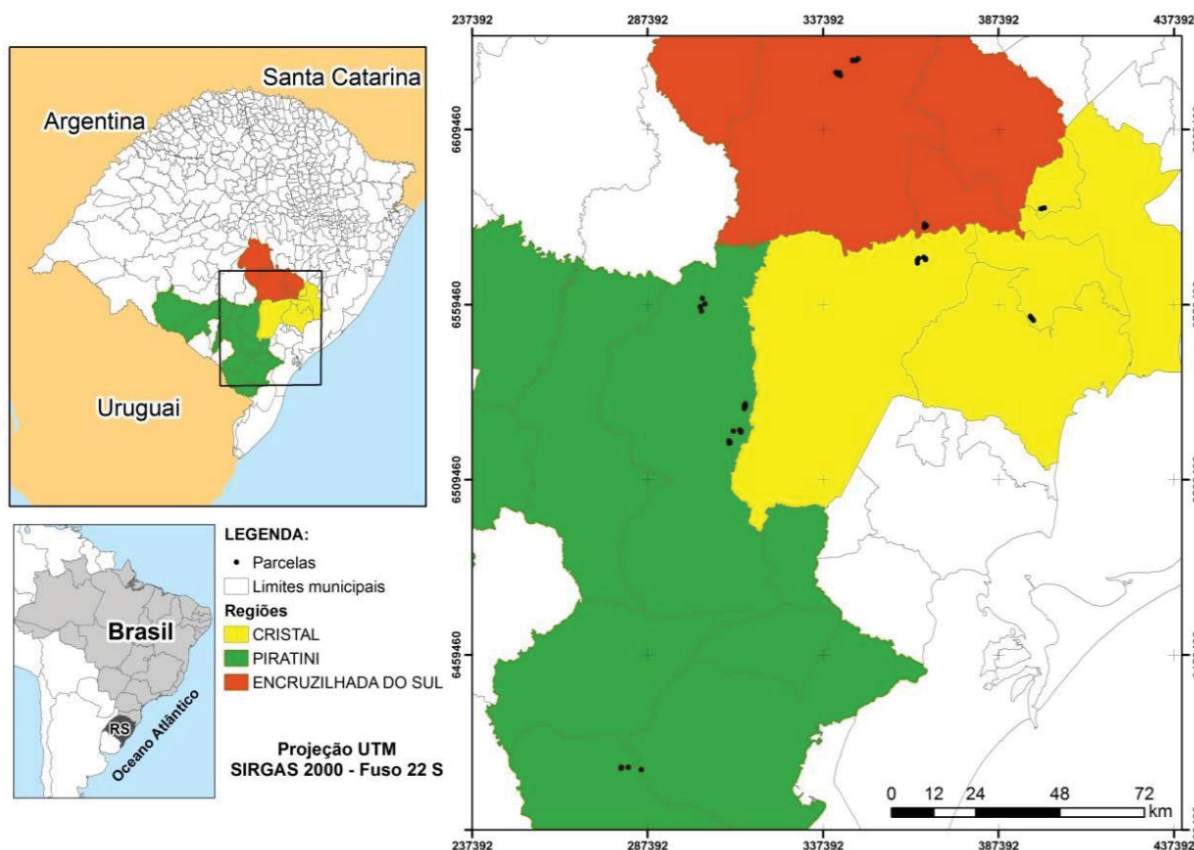
Salienta-se também que, normalmente na modelagem por aprendizagem de máquina, as bases de dados são divididas nos *sets*: treinamento e validação (GOLLAPUDI, 2016). Wilkins e Bruce (1987) descreveram que, na fase de treinamento, são utilizados dados para treinar modelos ao emparelhar a entrada fornecida com a saída esperada, sendo assim, o resultado dessa fase é o próprio modelo de aprendizado. O autor ainda afirma que a validação é a fase que visa medir a qualidade do modelo de aprendizado treinado e estimar as propriedades do modelo, como medidas de erro, precisão, entre outras. Essa fase usa um conjunto de dados de validação e a saída é um modelo de aprendizado sofisticado. Assim, nesta pesquisa, optou-se em realizar a modelagem valendo-se de treinamento e validação, visando manter as condições da modelagem da técnica testemunha.

### 3.2.1 Base de dados 670/2014 para treinamento dos algoritmos

Para o treinamento dos modelos de aprendizagem de máquina foi utilizado um conjunto de dados independentes ao utilizado para o treinamento. A independência foi garantida ao se utilizar uma nova base de dados e isso permitiu a generalização.

Foram amostrados povoamentos no estado do Rio Grande do Sul nas regiões dos municípios de Cristal, Encruzilhada do Sul e Piratini (Figura 8), sendo elas as regiões de maior concentração de plantio da espécie.

FIGURA 8 – ÁREA DE CULTIVO DE ACÁCIA-NEGRA NAS REGIÕES DE CRISTAL, ENCRUZILHADA DO SUL E PIRATINI, NO ESTADO DO RIO GRANDE DO SUL, BRASIL.



FONTE: Martins et al. (2019).

A amostragem abrangeu doze povoamentos e 670 árvores, com variação entre 1 a 10,75 anos. Em cada povoamento, foram instaladas, de modo aleatório, quatro parcelas<sup>5</sup> de forma circular com dimensão de 78,54 m<sup>2</sup> a fim de viabilizar a coleta das variáveis: biomassa total acima do solo, circunferência à altura do peito e altura total. A biomassa foi mensurada por meio da pesagem da árvore e de amostras para obtenção da biomassa seca, a altura com trena e a circunferência à altura do peito com uma fita métrica, conforme detalhado em Behling (2016).

Originalmente, para realização da modelagem da biomassa por regressão essa base de dados foi estratificada, sendo os dados agrupados em função da idade qualitativa dos povoamentos, de modo que foram definidas quatro classes de idades, conforme Tabela 1.

TABELA 1 – CLASSES DE IDADE DOS POVOAMENTOS DE ACÁCIA-NEGRA AMOSTRADOS NO ESTADO DO RIO GRANDE DO SUL, BRASIL.

Classe de idade	Idade	Local	<i>n</i>
-----------------	-------	-------	----------

<sup>5</sup>Parcela é a unidade na área de estudo (plantio florestal ou floresta) onde são coletados todas as variáveis necessárias (variáveis dendrométricas, solo, precipitação, ...) para o inventário florestal.

Muito Jovem	1,00	Piratini	71
Jovem	1,75	Cristal	59
	1,83	Encruzilhada do Sul	56
Média Inicial	2,33	Piratini	48
	2,75	Cristal	47
	3,08	Encruzilhada do Sul	68
Média Avançada	5,00	Piratini	43
	5,25	Cristal	60
	5,75	Encruzilhada do Sul	50
Madura	9,83	Piratini	51
	10,08	Cristal	57
	10,75	Encruzilhada do Sul	62

FONTE: O autor (2020).

LEGENDA:  $n$  é o número de árvores.

A classe Muito Jovem, proposta por Behling et al. (2019), foi agrupada na classe Jovem para análise exploratória dos dados.

A relação alométrica da biomassa em função da altura e diâmetro varia de acordo com a idade do povoamento e é afetada pelas classes diamétricas e de alturas, motivo pelo qual Behling et al. (2018) propôs as equações de biomassa para as quatro classes de idade mencionadas na Tabela 1. Dado que as idades variaram de 1 a 10,75 anos, a melhor divisão desses estratos não deveria incluir tamanhos de amplitude maior que dois anos, porque, dessa maneira, não seriam evidenciados os efeitos de tais proporções.

Na modelagem da biomassa por aprendizagem de máquina, ao invés de fazê-la por estrato, optou-se por incluir adicionalmente a variável independente idade, pois conforme definido na primeira hipótese desta pesquisa, as técnicas de aprendizagem de máquina são mais flexíveis e possuem maior capacidade de extrair e generalizar informações de uma base de dados, quando comparados aos modelos de regressão.

### 3.2.2 Base de dados 98/2012 para validação do ajuste dos algoritmos

Os dados utilizados para validação foram oriundos das regiões de Cristal e Piratini. O método para quantificação da biomassa é similar ao da primeira base, sendo detalhado em Behling (2014). O estudo compreendeu a amostragem de 8 povoamentos (4 em cada local de estudo), a instalação de 24 parcelas (3 em cada povoamento), nas quais foram amostradas 98 árvores. Os tamanhos das parcelas foram de 9 m x 16 m para os povoamentos com um ano de idade e 9 m x 14 m para



as demais idades, equivalendo a quatro linhas de plantio e 10 plantas em cada linha, totalizando 80 árvores. Mensurou-se, nas parcelas, a circunferência à altura do peito utilizando-se fita métrica graduada e as árvores para avaliação da biomassa foram selecionadas de maneira proporcional à quantidade árvores ao longo da distribuição de diâmetro obtida. O número de árvores amostradas ( $n$ ) por classe de idade (definida na Tabela 1) está destacado na Tabela 2.

TABELA 2 – NÚMERO DE ÁRVORES AMOSTRADAS ( $n$ ) PARA A BASE DE DADOS 98/2012 DE ACÁCIA-NEGRA NO ESTADO DO RIO GRANDE DO SUL, BRASIL.

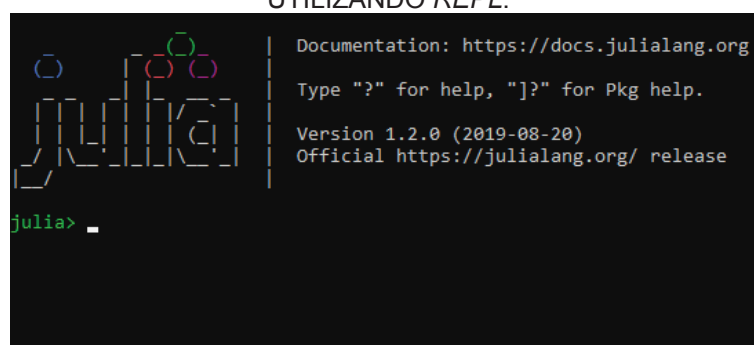
Classe	Idade	$n$
Jovem	2,0	22
Média Inicial	3,0-4,0	36
Média Avançada	5,0-6,0	40

FONTE: O autor (2020).

### 3.3 DOWNLOAD, INSTALAÇÃO DA LINGUAGEM DE PROGRAMAÇÃO JULIA E CARREGAMENTO DA BASE DE DADOS

O *download* da linguagem de programação *Julia* é realizado no *site* oficial do mesmo, vinculado ao link: <http://julialang.org>. Existem várias possibilidades de interface, como por exemplo: *Juno*, *Vs Code*, *Jupyter*, *JetBrains*, *Vim*, *Emacs*, *SublimeText*, *NotePad++*, *Julia Studio*, *REPL* (Read Evaluate Print Loop) *command line interface*, entre outras. Neste trabalho, optou-se pelo processamento de dados utilizando *REPL* (Figura 9).

FIGURA 9 – REPRESENTAÇÃO DA INTERFACE DA LINGUAGEM DE PROGRAMAÇÃO JULIA UTILIZANDO REPL.



FONTE: O autor (2020).

A importação da base de dados para a linguagem de programação *Julia* foi realizada conforme Algoritmo 1, em que a definição das variáveis da base de dados é apresentada na Figura 10. O caractere “#” é utilizado para adicionar comentários ao longo do código.

## ALGORITMO1 – IMPORTAÇÃO DE BASE DE DADOS NA LINGUAGEM DE PROGRAMAÇÃO JULIA

```
julia> # Importar a base de dados
julia> using DataFrames # a expressão using é utilizada para acionar o pacote
                        desejado
julia> dados=readtable("C:/Users/Usuário/OneDrive/Documentos/Mestrado/Dados/Dad
os Acacia/Dados/Dados.txt", separator = ' ')# dados é o nome do data frame
criado; readtable é expressão para ler um arquivo em formato txt; entre
""foi adicionada o local do armazenamento do arquivo txt no computador;
separator é a expressão utilizada para definir como os dados dentro do
arquivo estão separados, sendo que nesse caso eles foram separados por um
espaço e nesse caso representado por ' '.
```

FONTE: O autor (2020).

FIGURA 10 – BASE DE DADOS 670/2014 CARREGADA NO AMBIENTE DA LINGUAGEM DE PROGRAMAÇÃO JULIA



Row	_local String®	classe String®	i Float64®	d Float64®	h Float64®	b Float64®
1	Piratini	1 - Jovem	1.0	2.71	3.7	1.76
2	Piratini	1 - Jovem	1.0	2.23	3.0	1.14
3	Piratini	1 - Jovem	1.0	2.23	3.4	1.16
4	Piratini	1 - Jovem	1.0	2.55	2.9	1.52
5	Piratini	1 - Jovem	1.0	1.75	2.8	1.15
6	Piratini	1 - Jovem	1.0	2.39	2.9	1.56
7	Piratini	1 - Jovem	1.0	3.02	3.9	1.96
8	Piratini	1 - Jovem	1.0	2.86	3.6	2.49
9	Piratini	1 - Jovem	1.0	1.91	3.1	1.77
10	Piratini	1 - Jovem	1.0	1.59	2.6	1.17
11	Piratini	1 - Jovem	1.0	2.07	3.0	1.01
12	Piratini	1 - Jovem	1.0	2.55	3.5	1.78
13	Piratini	1 - Jovem	1.0	2.86	3.4	2.27
14	Piratini	1 - Jovem	1.0	2.39	3.32	1.36
15	Piratini	1 - Jovem	1.0	2.86	4.0	2.12
16	Piratini	1 - Jovem	1.0	1.59	2.55	1.29
17	Piratini	1 - Jovem	1.0	3.02	4.1	2.29
18	Piratini	1 - Jovem	1.0	3.02	4.2	2.31
653	Encruzilhada	4 - Madura	10.75	18.3	19.5	179.41
654	Encruzilhada	4 - Madura	10.75	10.66	13.2	34.47
655	Encruzilhada	4 - Madura	10.75	12.73	15.5	64.64
656	Encruzilhada	4 - Madura	10.75	12.57	15.8	73.0
657	Encruzilhada	4 - Madura	10.75	10.66	15.7	55.47
658	Encruzilhada	4 - Madura	10.75	12.25	16.5	77.45
659	Encruzilhada	4 - Madura	10.75	13.37	18.9	99.32
660	Encruzilhada	4 - Madura	10.75	13.05	18.0	81.83
661	Encruzilhada	4 - Madura	10.75	8.91	15.6	32.43
662	Encruzilhada	4 - Madura	10.75	12.41	16.3	83.35
663	Encruzilhada	4 - Madura	10.75	8.75	13.6	31.4
664	Encruzilhada	4 - Madura	10.75	11.3	16.5	56.34
665	Encruzilhada	4 - Madura	10.75	12.41	18.0	85.41
666	Encruzilhada	4 - Madura	10.75	3.82	7.7	3.89
667	Encruzilhada	4 - Madura	10.75	9.55	14.2	38.76
668	Encruzilhada	4 - Madura	10.75	10.66	15.4	39.91
669	Encruzilhada	4 - Madura	10.75	7.32	11.5	17.45
670	Encruzilhada	4 - Madura	10.75	14.32	17.8	109.61

FONTE: O autor (2020).

LEGENDA: *local* = região de origem da árvore; *classe* = classe de idade dos indivíduos; *i* = idade das árvores; *d* = diâmetro à altura do peito; *h* = altura total; *b* = biomassa total;

### 3.4 ANÁLISE EXPLORATÓRIA DAS BASES DE DADOS

A análise exploratória dos dados compreendeu o seguinte processamento:

#### i) Estatística descritiva:

Para as variáveis diâmetro à altura do peito, altura total e biomassa foram obtidas as estatísticas: média, variância, desvio-padrão, coeficiente de variação, erro padrão, erro absoluto, erro relativo, valor mínimo, valor máximo e intervalo de confiança para cada uma das bases de dados. Para demonstrar a variação dos valores das variáveis entre as classes de idade definidas, foram usados os gráficos do tipo *box-plot* e *violin*. Os comandos utilizados e as bibliotecas gráficas foram especificados na seção 4 e o processamento foi realizado para as duas bases de dados.

## ii) Correlações e definição das variáveis independentes para modelagem da biomassa

Foi analisado a correlação das variáveis biomassa, diâmetro à altura do peito, altura e idade por meio de gráficos, conforme apresentado na seção 4. Essa análise permitiu identificar as variáveis independentes a serem utilizadas, bem como a forma da relação dessas variáveis com a biomassa.

## iii) Escala da variável biomassa na modelagem

A variável dependente biomassa pode ser modelada em sua escala natural, logarítmica ou padronizada (intervalo de 0 a 1), por exemplo. Nesta seção, foram avaliadas evidências sobre qual escala da variável a modelagem deveria ser feita.

## iv) Relação alométrica da biomassa com as variáveis independentes diâmetro à altura do peito e altura para as bases de dados 670/2014 e 98/2012

Foram observadas as relações alométricas da variável biomassa com as variáveis diâmetro à altura do peito e altura total para as bases de dados de treinamento e validação, visando avaliar se essas relações alométricas foram semelhantes entre as bases de dados.

# 3.5 TÉCNICAS DE APRENDIZAGEM DE MÁQUINA USANDO A LINGUAGEM DE PROGRAMAÇÃO *JULIA*

A linguagem de programação *Julia* oferece uma variedade de recursos (bibliotecas e *packages*) para modelagem usando técnicas de aprendizagem de máquina. Os *packages* utilizados para o desenvolvimento da modelagem estão mencionados na Tabela 3.

TABELA 3 – *PACKAGES* UTILIZADOS NA MODELAGEM DA BIOMASSA UTILIZANDO A LINGUAGEM DE PROGRAMAÇÃO *JULIA*.

<i>Packages</i>	Biblioteca original
-----------------	---------------------

<i>DataFrames</i>	
<i>Statistics</i>	
<i>Distributions</i>	
<i>Random</i>	Linguagem <i>Julia</i>
<i>DecisionTree</i>	
<i>StatsPlots</i>	
<i>XLSX</i>	
<i>ScikitLearn</i>	
<i>PyCall</i>	
<i>PyPlot</i>	<i>Python</i>
<i>GadFly</i>	

FONTE: O autor (2020).

No algoritmo 2, é mostrado como recursos de outras linguagens de programação (neste caso *Python*) podem ser importados para dentro da linguagem *Julia*.

ALGORITMO2 – IMPORTAÇÃO DE *PACKAGES* NA LINGUAGEM DE PROGRAMAÇÃO *JULIA*.

```
julia> # Importar packages
julia> using Pkg # Pkg é o package responsável por importar ou clonar packages.
julia> Pkg.add("DataFrames") # .add define o comando de adicionar um package e “
” são utilizados para descrever que uma expressão é escrita e não
representa um Array ou um valor numérico.
julia> Pkg.clone("https://github.com/JuliaData/DataFrames.jl") # .clone define
quando clona-se um package de um site específico.
julia> Pkg.build("DataFrames") # .build é utilizado para reconstruir o package
atualizando o mesmo para novas versões.
```

FONTE: O autor (2020).

### 3.5.1 Técnicas de aprendizagem de máquina

Todos os algoritmos e extensões para aprendizagem de máquina se basearam no *package* principal *ScikitLearn.jl*. *ScikitLearn.jl* é a versão em linguagem de programação *Julia* da popular biblioteca *Scikit-learn*, sendo esse o motivo da escolha dessa biblioteca. Essa biblioteca possui muitos recursos, como por exemplo, suporte para *DataFrames*, ajuste de hiperparâmetros, uniões de recursos e *pipelines*, validação cruzada, entre outros.

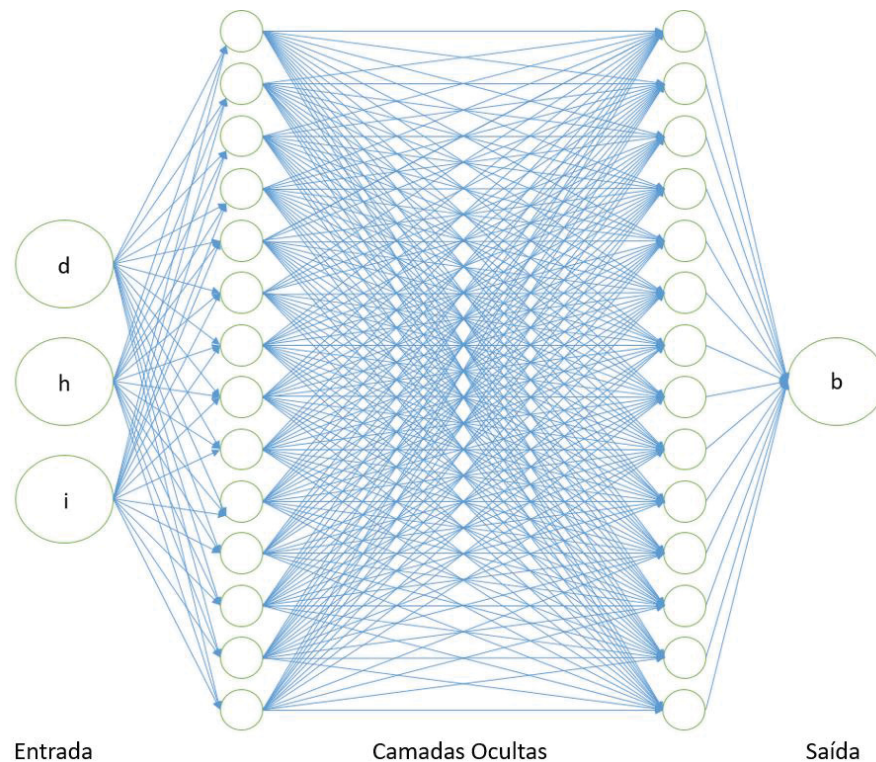
Os parâmetros dos algoritmos utilizados na criação das rotinas empregadas no presente trabalho estão descritos na sequência. Ressalta-se que todos os

parâmetros foram modificados em função da literatura, visto que não é objetivo deste trabalho avaliar o efeito de todas as possibilidades de modificação dos algoritmos. Para tanto, a principal literatura consultada foi Costa Filho (2019), que pesquisou quais parâmetros dos algoritmos influenciavam mais a acuracidade da modelagem de variáveis dendrométricas.

#### 3.5.1.1 Redes neurais artificiais

O algoritmo utilizado para a rede neural artificial foi o *backpropagation*. Utilizaram-se duas camadas ocultas com quinze neurônios em cada, usando a função de ativação logística. O número de épocas foi fixado no valor máximo (*max iter*) de 1.000. Optou-se por variar o tipo de *solver* empregado utilizando-se o *LBFGS*, por ser um otimizador da família dos métodos *quasi-Newton*, sendo empregada em bases de dados inferiores a 1.000 repetições. As variáveis que apresentaram alterações foram o termo *momentum*, variando de 0,3 a 0,6; e a taxa de aprendizado (*learning rate init*), que variou de 0,01 a 0,30. A estrutura da rede é ilustrada na Figura 11.

FIGURA 11 – REDE NEURAL ARTIFICIAL UTILIZADA NA PREDIÇÃO DOS DADOS 670/2014



FONTE: O autor (2020).

LEGENDA:  $i$  = idade das árvores;  $d$  = diâmetro à altura do peito;  $h$  = altura total;  $b$  = biomassa total;

### 3.5.1.2 $k$ -vizinhos mais próximos

O treinamento dos regressores  $k$ -vizinhos mais próximos considerou o número de vizinhos variando de 1 a 30 e de 30 a 150. As métricas de distâncias empregadas foram: *Euclidiana*, *Manhattan* e *Chebyshev*. Um total de 324 combinações foram avaliadas.

### 3.5.1.3 Florestas aleatórias

O treinamento da técnica de florestas aleatórias foi realizado por meio do método *bagging*, consistindo na utilização de subconjuntos de dados da base de treinamento para geração das árvores; o número de árvores de regressão ( $n\_estimators$ ) e o número de amostras selecionadas em cada nó ( $min\_samples\_leaf$ ), que consistiram nos parâmetros que variaram no treinamento dos modelos. Testou-se valores de 10 a 100, com saltos de 10 unidades a cada treinamento, de 100 a 500 foram testados com saltos de 50 unidades. Já para o número de variáveis selecionadas em cada nó foram testados os valores 1, sendo o valor padrão do pacote, 2 e 3. Por fim, o parâmetro *out of bag score* ( $oob\_score$ ) foi testado com

ativado (*true*) e desativado (*false*). Dessa forma, um total de 108 combinações foram testadas.

### 3.5.1.4 Máquinas de vetores de suporte

As máquinas de vetores de suporte foram configuradas com função *kernel de base radial* (*radial base function*), dessa forma, tem-se o parâmetro  $\gamma$  (*gamma*). Os valores de  $\gamma$  variaram de  $2^{-8}$  a  $2^{-1}$ . O parâmetro  $C$  foi testado com uma variação de  $2^1$  a  $2^8$ , sendo que o intervalo para ambos é de  $2^1$ . Para *épsilon* utilizou-se os valores 0,10, padrão do algoritmo e 0,50 proposto pelo autor. Sendo assim, o número de modelos testados foi de 128.

## 3.6 EQUAÇÕES DE REGRESSÃO DE BIOMASSA – TÉCNICA TESTEMUNHA

Visto que a modelagem por regressão é o método mais utilizado, para predição de biomassa, em geral, optou-se por comparar os resultados obtidos com a melhor técnica de aprendizagem de máquina obtida com o método de regressão. Dessa forma, utilizou-se as equações não lineares propostas por Behling et al. (2019), que foram ajustadas utilizando-se a base de dados 670/2014. A modelagem foi realizada por estrato de idade, definido na Tabela 1. As equações foram ajustadas utilizando-se o procedimento **Weighted-Nonlinear Seemingly Unrelated Regression** (**WNSUR**), conforme detalhado em Behling et al. (2019). As equações ajustadas estão apresentadas na Tabela 4.

TABELA 4 – EQUAÇÕES DE BIOMASSA TOTAL ACIMA DO SOLO PARA ÁRVORES DE ACÁCIA-NEGRA ORIUNDAS DE POVOAMENTOS NO ESTADO DO RIO GRANDE DO SUL, BRASIL.

Classe	Modelos
Muito Jovem	$b = 0,205422(d^2h)^{0,459328} + 0,17493(d^2h)^{0,506569}$
Jovem	$b = 0,0324844(d^2h)^{0,827296} + 0,052046(d^2h)^{0,837992}$
Média Inicial	$b = 0,003320(d^2h)^{1,079325} + 0,039548(d^2h)^{0,899084}$
Média Avançada	$b = 0,003976(d^2h)^{0,863215} + 0,051517d^{1,979083}h^{0,692747}$
Madura	$b = 0,001216(d^2h)^{1,165186} + 0,053628d^{1,967389}h^{0,736520}$

FONTE: Adaptado de Behling et al. (2019).

LEGENDA:  $b$  é biomassa (kg);  $d$  é diâmetro à altura do peito (cm);  $h$  é altura total (m);

### 3.7 CRITÉRIOS DE AVALIAÇÃO DA QUALIDADE PREDITIVA DA BIOMASSA POR TÉCNICAS DE APRENDIZAGEM DE MÁQUINA E MODELOS DE REGRESSÃO

Para avaliar a qualidade preditiva da biomassa pelas técnicas e algoritmos de aprendizagem de máquina optou-se pelo uso das estatísticas de viés, precisão e acurácia. Essas foram calculadas utilizando-se a base de dados de validação 98/2012, mas também foram calculadas para a base de treinamento.

O viés, ou *bias*, é uma suposição adotada pelos algoritmos de forma a generalizar os dados de treinamento. Na estatística tem-se dois tipos de viés, o absoluto (eq. 10) e o relativo (eq. 11), que indicam o erro sistemático associado a análise para uma dada amostra.

Mikhail e Achermann (1976) definem acurácia (eqs. 14 e 15) como sendo o grau de proximidade de uma estimativa com seu parâmetro, enquanto precisão expressa o grau de consistência da grandeza medida com sua média. Os autores acrescentam que acurácia reflete a proximidade de uma grandeza estatística ao valor do parâmetro para o qual foi estimada e que precisão está diretamente ligada com a dispersão da distribuição das observações.

De modo geral, o viés e a precisão (eqs. 12 e 13) combinados definem a performance de um estimador, sendo que quanto maior o viés, menos preciso é o estimador e menor é sua capacidade de estimar acuradamente um ponto (WALTHER; MOORE, 2005). Acurácia é definida como a distância entre valores estimados e os valores reais (BAINBRIGE, 1985; JONES, 1997; KREBS, 2007; ZAR, 2009). Salienta-se que precisão alta (valores pequenos dessa estatística) e viés pequenos levam a estimadores altamente acurados, contudo, um estimador inacurado pode ser devido a uma alta variância das variáveis (baixa precisão) ou a vieses elevados.

$$\bar{e} = \frac{\sum_{i=1}^n e_i}{n} = \frac{\sum_{i=1}^n (b - b_{est})}{n} \quad (10)$$

$$\bar{e}_{\%} = \frac{\bar{e}}{\bar{b}} 100 \quad (11)$$

$$S_e = \sqrt{\frac{\sum_{i=1}^n (b - \bar{e} - b_{est})^2}{n - 1}} = \sqrt{\frac{\sum_{i=1}^n (e_i - \bar{e})^2}{n - 1}} \quad (12)$$

$$S_{e\%} = \frac{S_e}{\bar{b}} 100 \quad (13)$$

$$m_x = \sqrt{(S_e)^2 + (\bar{e})^2} \quad (14)$$

$$m_{x\%} = \frac{m_x}{\bar{b}} 100 \quad (15)$$

Em que:  $\bar{e}$  é o viés absoluto,  $\bar{e}_{\%}$  é o viés relativo,  $e_i$  é o erro sistemático para i-ésima árvore,  $b$  é a biomassa observada,  $b_{est}$  é a biomassa estimada,  $\bar{b}$  é a média da biomassa observada,  $n$  é o número de árvores,  $S_e$  é a precisão absoluta,  $S_{e\%}$  é a precisão relativa,  $m_x$  é a acurácia absoluta e  $m_{x\%}$  é a acurácia relativa.

Para facilitar a visualização dos resultados, serão colocados os resultados de viés precisão e acurácia dos vinte melhores resultados.

Com o intuito de auxiliar na visualização dos estimadores optou-se pelo uso de gráficos de dispersão da biomassa observada e estimada em função da variável diâmetro à altura do peito, pois é uma das relações alométricas mais conhecidas (modelo de potência). Para cada técnica de aprendizagem de máquina, foram selecionados os quatro melhores estimadores, definidos pela avaliação das estatísticas definidas anteriormente. Adicionalmente, foi realizada a avaliação gráfica da dispersão dos resíduos em função da variável diâmetro à altura do peito.



## 4 RESULTADO E DISCUSSÃO

### 4.1 ANÁLISE EXPLORATÓRIA DAS BASES DE DADOS

#### 4.1.1 Estatística descritiva

A programação utilizada para obtenção das estatísticas valor mínimo, valor máximo, média, variância, desvio-padrão, coeficiente de variação, erro-padrão e intervalo de confiança está descrita no Algoritmo 3. Para elaborar a função utilizou-se os *packages*: *Statistics* e *Distributions*.

ALGORITMO3 – CÁLCULO DA ESTATÍSTICA DESCRITIVA NA LINGUAGEM DE PROGRAMAÇÃO JULIA.

```
julia> # Estatística Descritiva
julia> variavel = dados.d; # definição da variável que será feita a estatística
descritiva, nesse exemplo utilizou-se a variável d (diâmetro à altura do
peito). Vale ressaltar que dados é a base de dados carregada conforme
Figura 9.
julia> function ED(variavel)# inicialização da função para a base de dados
denominada de dados.
    n = length(variavel) # número de árvores da amostra.
    min = minimum(variavel) # valor mínimo.
    max = maximum(variavel) # valor máximo.
    x = mean(variavel) # média.
    var = var(variavel) # variância.
    dp = std(variavel)# desvio padrão.
    CV = dp/x*100# coeficiente de variação, em porcentagem.
    Sx = dp/sqrt(n)# erro padrão.
    conf_level = 0.95# nível de confiança escolhido para a estatística
    alpha = (1-conf_level)# probabilidade do erro tipo I.
    t = quantile(TDist(n-1), 1-alpha/2)# valor da estatística t de Student.
    ICi = x - Ea # limite do intervalo de confiança inferior
    ICs = x + Ea # limite do intervalo de confiança superior
    println("Valor mínimo: ", min) # println é a expressão utilizada para
imprimir a frase desejada, nesse caso um texto expresso entre"". Após a
virgula declara-se o valor que será imprimido logo após a frase criada.
    println("Valor máximo: ", max)
    println("Média: ", x)
    println("Variância: ", var)
```

```
println("Desvio Padrão: ", dp)
println("Coeficiente de Variação (%):", CV)
println("Erro padrão: ", Sx)
println("IC inferior: ", ICi)
println("IC superior: ", ICs)
end# expressão para finalizar funções
julia> # Empregar a função
julia> ED(variavel)
```

FONTE: O autor (2020).

Na Tabela 5 estão apresentados os resultados para as estatísticas descritivas das variáveis diâmetro à altura do peito (cm), altura total (m) e biomassa total (kg) para a base de dados 670/2014. As variáveis demonstram uma variação acentuada, em função da variação na idade dos plantios, sendo de 1 a 10,75 anos, especialmente para a biomassa.

TABELA 5 – ESTATÍSTICA DESCRITIVA PARA AS VARIÁVEIS DIÂMETRO À ALTURA DO PEITO (cm), ALTURA TOTAL (m) E BIOMASSA TOTAL (kg) PARA A BASE DE DADOS DE **TREINAMENTO 670/2014**,

Estatística descritiva	<i>d</i> (cm)	<i>h</i> (m)	<i>b</i> (Kg)
Valor mínimo	0,6400	1,6000	0,3114
Valor máximo	23,5500	21,9000	361,0660
Média	9,5019	12,1704	45,5242
Variância	19,2838	24,4544	2.315,9604
Desvio Padrão	4,3913	4,9451	48,1244
Coeficiente de Variação (%)	46,2153	40,6327	105,7117
Erro padrão	0,1695	0,1909	1,8578
IC inferior	9,1690	9,8767	41,8764
IC superior	9,8348	11,7955	49,1721

FONTE: O autor (2020).

LEGENDA: *d* é diâmetro à altura do peito; *h* é altura total; *bt* é a biomassa total; *IC* é intervalo de confiança.

Os intervalos das variáveis independentes para a base de dados de validação98/2012 são apresentados na Tabela 6.

TABELA 6 – ESTATÍSTICA DESCRITIVA PARA AS VARIÁVEIS DIÂMETRO À ALTURA DO PEITO (cm), ALTURA TOTAL (m) E BIOMASSA TOTAL (Kg) PARA A BASE DE DADOS DE VALIDAÇÃO 98/2012.

Estatística descritiva	<i>d</i> (cm)	<i>h</i> (m)	<i>b</i> (Kg)
Valor mínimo	3,82	6,10	3,73
Valor máximo	21,49	21,60	273,17
Média	11,17	14,47	74,86
Variância	22,06	18,66	4742,22
Desvio Padrão	4,70	4,32	68,86
Coeficiente de Variação (%)	42,07	29,86	91,99
Erro padrão	0,53	0,49	7,80
IC inferior	10,11	12,14	59,33
IC superior	12,23	13,49	90,39

FONTE: O autor (2020).

LEGENDA: *d* é diâmetro à altura do peito; *h* é altura total; *bt* é a biomassa total; *IC* é intervalo de confiança.

Observou-se que os valores mínimos e máximos da base de dados 98/2012 estão dentro do intervalo de valores mínimos e máximos da base de dados 670/2014, o que corrobora para o uso desses dados para a validação visto que estimar dados fora do intervalo causaria problemas nas estimativas, reduzindo a qualidade dos resultados.

Na criação do gráfico composto por *violinplot*, *boxplot* e *dotplot* (Algoritmo 4), utilizou-se o *package* StatsPlots.

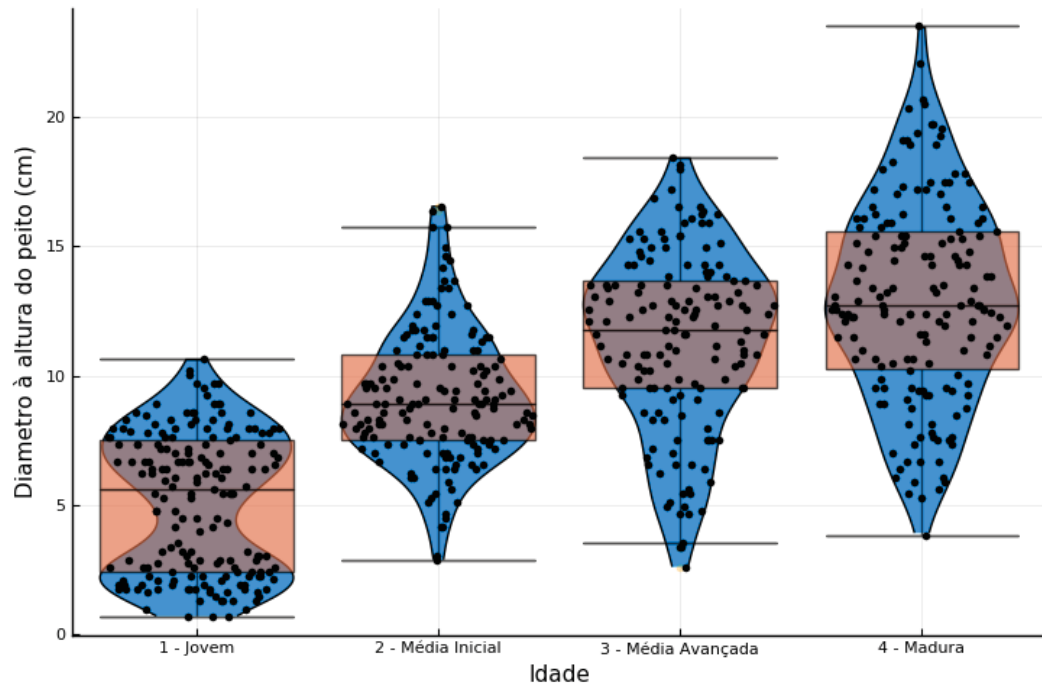
ALGORITMO4 – PROGRAMAÇÃO PARA CRIAÇÃO DOS GRÁFICOS: VIOLINPLOT, BOXPLOT E DOTPLOT NA LINGUAGEM DE PROGRAMAÇÃO JULIA.

```
julia> # Gráfico violinplot
julia> default(size = (900,900)) # dimensão do gráfico em pixels
julia> pyplot()
julia> @df dados violin(:classe, :d, xlabel = "Idade", ylabel = "Diâmetro à altura
do peito (cm) ", legend = false, marker = (0.2,:blue, stroke(0)), label =
"Violinplot") # false na expressão legend define que não haverá legenda
dessa forma não serão expressos os Labels, mas os mesmos foram adicionados
para demonstrar como introduzi-los; o valor 0.2 no marker define a dimensão
do marcador, em sequência tem-se a cor (azul nesse caso) e o stroke que é
a dimensão do ponto de fechamento do violinplot.
julia> @df dados boxplot!(:classe, :d, marker=(0.3, :orange, stroke(0)),
alpha=0.65, label= "Boxplot") # alpha define a opacidade do fundo da
caixa.
```

```
julia> @df dados dotplot!(:classe, :d, marker = (:black , stroke(0)), label =
    "Dotplot")
julia> # Armazenamento pode ser realizado dentro da plataforma PyPlot.
```

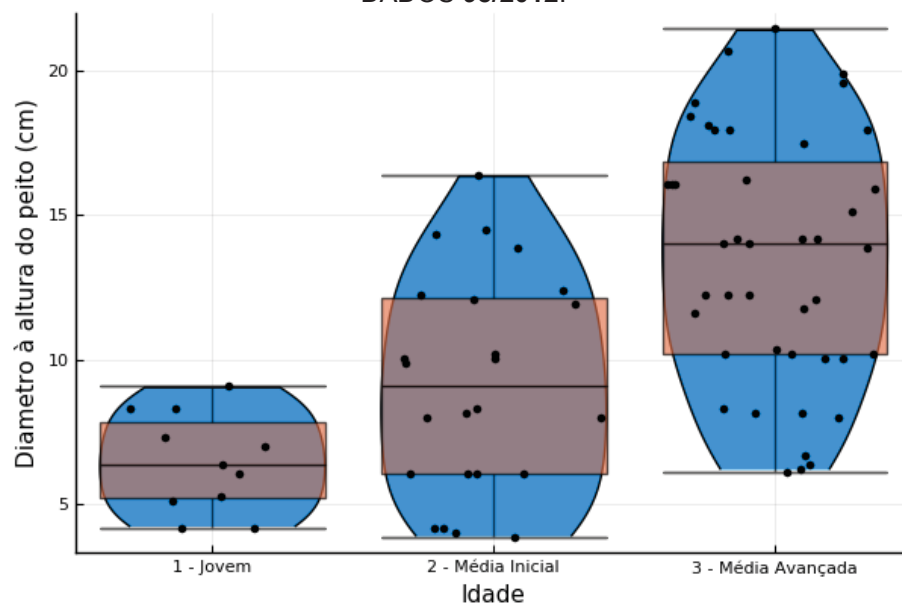
FONTE: O autor (2020).

FIGURA 12 – GRÁFICO *VIOLINPLOT*, *BOXPLOT* E *DOTPLOT* PARA A VARIÁVEL  $d$  (cm), PARA AS CLASSES DE IDADES: JOVEM, MÉDIA INICIAL, MÉDIA AVANÇADA E MADURA PARA A BASE DE DADOS 670/2014.



FONTE: O autor (2020).

FIGURA 13 – GRÁFICO *VIOLINPLOT*, *BOXPLOT* E *DOTPLOT* PARA A VARIÁVEL  $d$  (cm), PARA AS CLASSES DE IDADES: JOVEM, MÉDIA INICIAL E MÉDIA AVANÇADA PARA A BASE DE DADOS 98/2012.



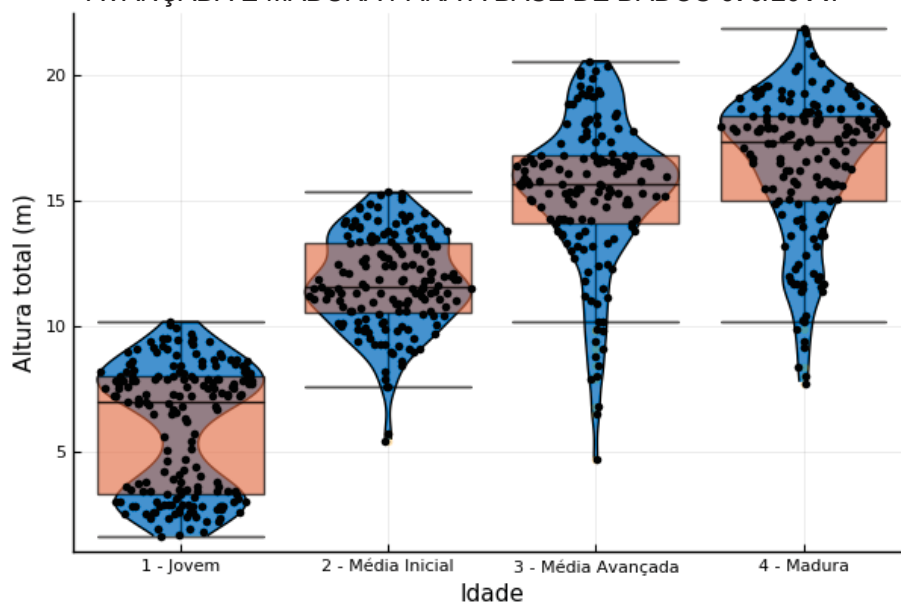
FONTE: O autor (2020).

Para a base de dados 670/2014, a variável diâmetro à altura do peito na idade Jovem teve uma condição de bimodalidade, devido a frequência distinta entre os valores da idade 1 ano para as demais (1,75 e 1,83 anos) (Figura 12). A simetria da variável mudou ao longo das idades, de modo que na idade madura foi observada uma condição de aproximação de simetria. Alguns valores demonstraram-se potenciais *outliers*, os quais foram verificados e considerados válidos devido a variabilidade dos dados. Devido ao crescimento da floresta, a amplitude aumentou conforme a idade. A variabilidade dos diâmetros da base de dados de treinamento (Figura 12) e validação (Figura 13) ocuparam o mesmo domínio, o que é apropriado para a modelagem.

Finger (1982), em seu estudo com acácia-negra de idades entre 3,5 a 7,5 anos, observou que sua distribuição de diâmetros era assimétrica à direita na idade de 3,5 anos e à esquerda nas demais idade. O autor ainda descreveu que ocorreu redução da curtose das distribuições diamétricas em todas as idades.

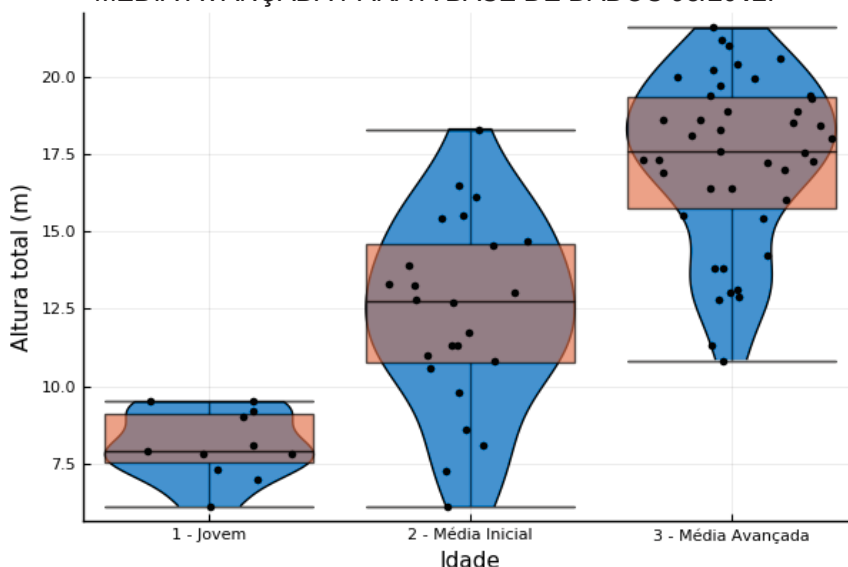
Realizou-se o mesmo processo com as variáveis altura total e biomassa, para a base de dados 670/2014 (Figura 14 e 16) e 98/2012 (Figura 15 e 17).

FIGURA 14 – GRÁFICO VIOLINPLOT, BOXPLOT E DOTPLOT PARA A VARIÁVEL ALTURA (m) DE ÁRVORES DE ACÁCIA NEGRA, PARA AS CLASSES DE IDADES: JOVEM, MÉDIA INICIAL, MÉDIA AVANÇADA E MADURA PARA A BASE DE DADOS 670/2014.



FONTE: O autor (2020).

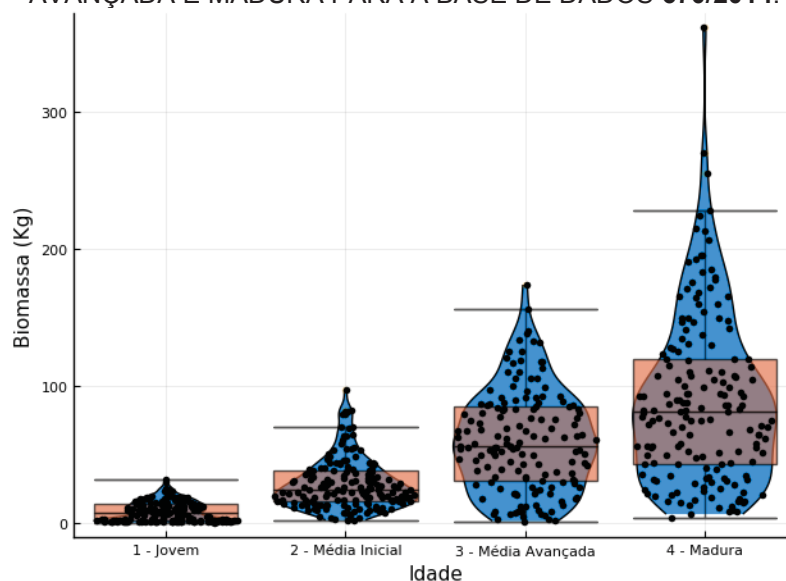
FIGURA 15 – GRÁFICO *VIOLINPLOT*, *BOXPLOT* E *DOTPLOT* PARA A VARIÁVEL ALTURA (m) DE ÁRVORES DE ACÁCIA NEGRA, PARA AS CLASSES DE IDADES: JOVEM, MÉDIA INICIAL E MÉDIA AVANÇADA PARA A BASE DE DADOS 98/2012.



FONTE: O autor (2020).

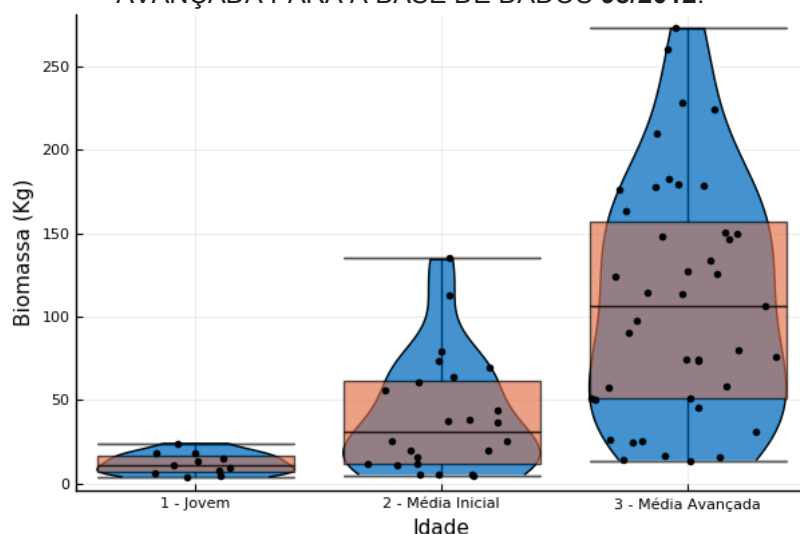
Para a base de dados 670/2014, a variável altura total na idade Jovem teve uma condição de bimodalidade (Figura 12). A simetria da variável também demonstrou mudanças ao longo das idades, contudo, nessa variável não se observou condições de aproximação de simetria. Aliás, a assimetria aumentou para os povoamentos mais velhos para ambas as bases de dados.

FIGURA 16 – GRÁFICO *VIOLINPLOT*, *BOXPLOT* E *DOTPLOT* PARA A VARIÁVEL BIOMASSA DE ACÁCIA NEGRA (Kg), PARA AS CLASSES DE IDADES: JOVEM, MÉDIA INICIAL, MÉDIA AVANÇADA E MADURA PARA A BASE DE DADOS 670/2014.



FONTE: O autor (2020).

FIGURA 17 – GRÁFICO VIOLINPLOT, BOXPLOT E DOTPLOT PARA A VARIÁVEL BIOMASSA DE ACÁCIA NEGRA (Kg), PARA AS CLASSES DE IDADES: JOVEM, MÉDIA INICIAL E MÉDIA AVANÇADA PARA A BASE DE DADOS 98/2012.



FONTE: O autor (2020).

Tratando-se da variável biomassa, observa-se uma mudança na forma ao comparar-se com as outras variáveis, isso ocorre em virtude da variável se comportar de maneira contrária as variáveis: diâmetro à altura do peito (cm) e altura total (m). Para a biomassa, a distribuição evoluiu dos povoamentos muito jovens para a idade madura de forma a obter uma distribuição assimétrica positiva, para altura negativa e para o diâmetro se aproximando da simetria. Dessa forma, com o avanço da idade da floresta, tem-se mais árvores com baixo estoque de biomassa, a maioria das árvores são altas e há um equilíbrio de tamanho em relação ao diâmetro. Assim, dada essas condições, é de se esperar que as relações alométricas da biomassa com as variáveis diâmetro à altura do peito e altura irão mudar ao longo das idades da floresta.

A condição de bimodalidade observada nas outras variáveis também foi observada para a biomassa no conjunto de dados muito jovem, devido a união das idades, conforme destacado anteriormente.

#### 4.1.2 Correlações e definição das variáveis independentes para modelagem da biomassa

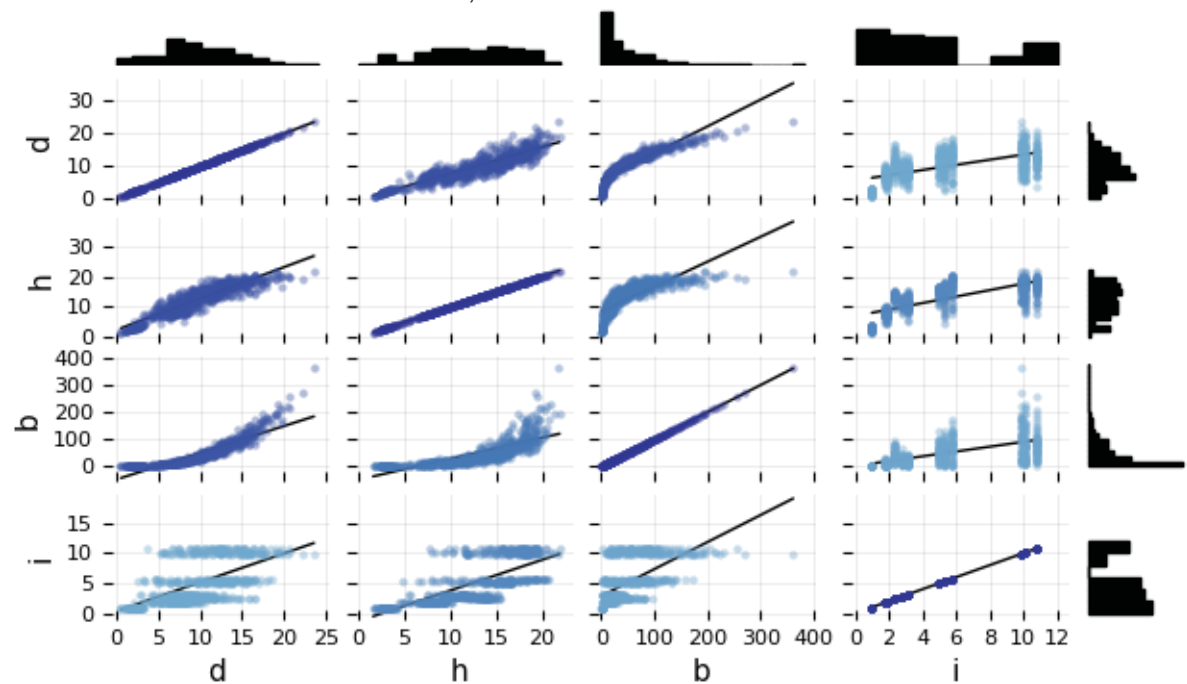
A correlação entre as variáveis, avaliada por meio de um gráfico *cornerplot*, está representada na Figura 18. Para construção desse gráfico, foram utilizados os pacotes *PyPlot* e *StatsPlots*, conforme Algoritmo 5.

ALGORITMO5 – REPRESENTAÇÃO DA PROGRAMAÇÃO DE *CORNERPLOT* EM LINGUAGEM DE PROGRAMAÇÃO JULIA.

```
julia> # Gráfico cornerplot
julia> gr(size = (900,900)) # dimensão do gráfico em pixels
julia> plotly()
julia> @df dados cornerplot([:d :h :b :idade], grid = false) # representa a criação
do cornerplot, em que entre [] tem-se as variáveis oriundas da base de
dados; o grid com a expressão false elimina as linhas dentro dos gráficos.
julia> # O armazenamento pode ser feito dentro da plataforma plotly.
```

FONTE: O autor (2020).

FIGURA 18 – GRÁFICO *CORNERPLOT* PARA AS VARIÁVEIS DIÂMETRO À ALTURA DO PEITO (*d*), EM cm, ALTURA TOTAL (*h*), EM m, BIOMASSA TOTAL (*b*), EM Kg E IDADE (*i*), EM anos, DE ACÁCIA NEGRA, PARA A BASE DE DADOS 670/2014.



FONTE: O autor (2020).

Em que: Quanto maior for a tonalidade de azul, a correlação se aproxima do valor 1.

Foi observado que a relação da biomassa com as variáveis diâmetro à altura do peito e altura é não linear, motivo pelo qual Behling et al. (2019) propuseram as



equações não lineares de biomassa. Adicionalmente, nota-se que esse dimensionamento está associado ao tema alometria e, conforme Niklas (1994) e Brown et al. (2004), segue um comportamento da lei de potência (*power-law*).

A biomassa é mais correlacionada com o diâmetro, seguido da variável altura e também com a idade, que por sua vez, se comportou como uma relação linear. As variáveis independentes (diâmetro à altura do peito, altura e idade) também são correlacionadas entre si, mas não são fortemente correlacionadas e isso é desejável para a modelagem.

Uma forte relação entre o diâmetro à altura do peito e a biomassa de árvore é esperada, de acordo com diversos trabalhos: Higuchi e Carvalho (1994), Nelson et al. (1999), Chaves et al. (2005) e Robortella (2010), Silveira (2009) e de Melo et al. (2014), os quais descrevem que a correlação da biomassa com o diâmetro à altura do peito foram altas para áreas de floresta ombrófila densa e caixeta, *Tabebuia cassinoides*, respectivamente. A altura não apresentou a mesma correlação, apresentando valores menores. Schikoski et al. (2013) descrevem, para *Pinus* spp., que as variáveis diâmetro à altura do peito, altura e idade possuem alta correlação com a biomassa, apresentando correlação de 0,916, 0,910 e 0,931, respectivamente. Essas literaturas demonstram que a base de dados apresenta tendências semelhantes com as três variáveis independentes sendo importantes para calcular a variável dependente, mesmo de espécies diferentes.

Embora a idade possua menor correlação com a biomassa, quando comparada com as outras duas variáveis independentes, ela é importante no sentido de uma estratificação na modelagem de aprendizagem de máquina, assim como pensado pelos autores que propuseram a modelagem via regressão. A razão pela qual a estratificação por idade foi implementada é que as relações alométricas das variáveis diâmetro à altura do peito e altura com a biomassa varia com a idade do povoamento, o que se torna claro ao analisar os coeficientes das regressões apresentados na Tabela 4. De acordo com Niklas (1995), a idade do povoamento influencia as relações entre a biomassa e as variáveis que expressam as dimensões da árvore, normalmente, diâmetro à altura do peito e altura.

Saint-André et al. (2005), por meio de sua pesquisa, afirmam que os coeficientes dos modelos de biomassa variaram com a idade do povoamento, sendo um resultado da mudança da maturidade das árvores combinado com sua posição sociológica. Os autores relatam que duas árvores com as mesmas dimensões das

variáveis diâmetro à altura do peito e altura, mas com diferentes idades, são diferentes quanto à sua ocupação na posição sociológica. A árvore mais jovem pode ser dominante e a mais velha provavelmente é suprimida, assim, ambos os efeitos (idade e posição sociológica) causam um aumento da biomassa do fuste para a árvore mais antiga, porque a densidade da madeira aumenta conforme a idade do povoamento e o fator de forma do fuste é maior para árvores dominadas. Essas condições também são observadas para as bases de dados utilizadas nesta pesquisa e revela a importância de utilizar a idade como um preditor, no sentido de estratificar as relações alométricas indiretamente por meio dos modelos de aprendizagem de máquina.

Dada as evidências de correlação, julgou-se apropriado na modelagem da biomassa por aprendizagem de máquina a incluir as variáveis diâmetro à altura do peito, altura e idade como variáveis independentes, que são variáveis acessíveis nos inventários florestais. Em algumas modelagens, as variáveis diâmetro à altura do peito e altura são transformadas ( $1/d$ ,  $d^2$ ,  $1/h$ ,  $1/h^2$ ) ou combinadas ( $dh$ ,  $d^2h$ ,  $1/d^2h$ ) com objetivo de aumentar a acuidade da modelagem, como por exemplo, nos trabalhos desenvolvidos por (ROLIM et al., 2006; AZEVEDO et al., 2011; DE MIRANDA et al., 2015; SILVA et al., 2016; DA SILVA et al., 2018; REGO et al., 2019), dessa forma, em muitos casos, houve um aumento na acuidade das estimativas. Neste trabalho isso não foi realizado pelos seguintes motivos:

i) Os modelos de regressão foram ajustados somente com as variáveis diâmetro à altura do peito e altura, sem transformação ou combinação, sendo o as equações propostas por idade. Assim, para manter as mesmas condições de comparação, seria apropriado manter a mesma estrutura das variáveis independentes.

ii) Os modelos de aprendizagem de máquina são flexíveis e eles devem ter a capacidade de assumir indiretamente as relações de combinação ou de transformação das variáveis independentes. Assim, a acurácia não poderia ser muito afetada pela combinação ou transformação dessas variáveis. Logicamente, qualquer variável adicional, mesmo que a correlação se aproxime de zero, ela contribui para reduzir a soma de quadrado de resíduos, conseqüentemente, aumentar a acuidade da predição.

Sanquetta et al. (2015), relatam que uma das formas de melhorar a qualidade da predição da biomassa por meio de técnicas de aprendizagem de máquina é a inclusão de outras variáveis preditoras no modelo, tais como: índice de sítio,

densidade, entre outras. Variáveis meteorológicas e de solo, inclusive disponíveis para as duas bases de dados 670/2014 e 98/2012, também podem ser inseridas como variáveis independentes para modelagem da biomassa. As relações alométricas são dependentes do meio em que as árvores estão e são afetadas por fatores que podem ser medidos indiretamente por meio de variáveis que caracterizam o meio, como as variáveis meteorológicas e de solo.

Poorter et al. (2012) discutem sobre como o padrão de alocação da biomassa dos componentes é influenciado pelo ambiente de crescimento, tamanho da planta, história evolutiva e da concorrência, consequentemente, refletindo na alometria. Apesar da notória importância da inclusão de variáveis do ambiente na modelagem, essas variáveis são mais caras de mensurar e demandam mais tempo que propriamente diâmetro à altura do peito, altura e idade, o que desmotiva o seu uso, principalmente em empresas florestais.

Não se utilizou outras variáveis nos ajustes dos estimadores para comparar diretamente as aprendizagens de máquina com o método de testemunha. Sendo assim, empregou-se as mesmas variáveis utilizadas nos ajustes dos regressores *WNSUR* (diâmetro à altura do peito, altura total e idade).

#### 4.1.2 Escala da variável biomassa na modelagem

Dada as evidências de relações observadas na Figura 10, a modelagem da biomassa foi em sua forma original, sem transformações, pelos seguintes motivos:

- As relações entre a variável biomassa com diâmetro à altura do peito e altura são não lineares. Elas seguem o comportamento da lei de potência.
- A linearização da variável biomassa resulta em um viés inerente à transformação utilizada nos dados em relação à escala aritmética original e sua correção é necessária. Na modelagem por regressão, a forma de realizar a correção da discrepância logarítmica é bem conhecida, mas para a modelagem por aprendizagem de máquina não é clara e não se tem certeza se a correção do viés é assegurada. Assim, a modelagem direta da variável biomassa permite trabalhar com os valores originais e não implica na transformação dos dados.
- Em uma simulação de padronização das variáveis (menor valor zero e maior valor de um), não foi observada uma melhora em relação a

performance do ajuste das técnicas e algoritmos de aprendizagem de máquina.

- Os modelos de aprendizagem de máquina são não paramétricos e possuem capacidade de modelar relações não lineares. Aliás, no início desta pesquisa, a modelagem havia sido realizada por meio da transformação logarítmica da biomassa, pois acreditava-se que os ajustes dos modelos de aprendizagem de máquina seriam mais rápidos e com menos viés na distribuição de resíduos, mais isso não foi evidenciado.

#### 4.1.3 Relação alométrica da biomassa com as variáveis independentes diâmetro à altura do peito e altura para as bases de dados 670/2014 e 98/2012

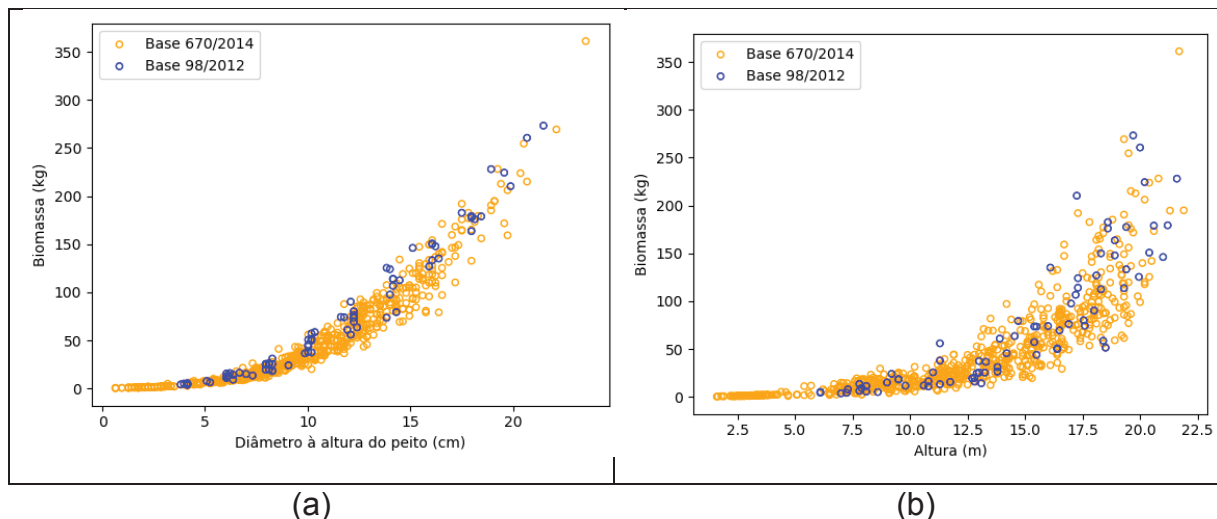
Para evidenciar o comportamento da relação da biomassa com as variáveis independentes diâmetro à altura do peito e altura foram construídos gráficos de dispersão, conforme Algoritmo 6. O algoritmo utilizou o *package PyPlot*. O resultado pode ser visualizado na Figura 19.

##### ALGORITMO6 – PROGRAMAÇÃO PARA CRIAÇÃO DOS GRÁFICOS DE DISPERSÃO PARA COMPARAÇÃO NA LINGUAGEM DE PROGRAMAÇÃO JULIA.

```
julia> # Gráfico de diâmetro à altura do peito
julia> scatter(dados.d, dados.b, facecolor= "none", edgecolors="orange", s=20,
label="Base 670/2014") # nuvem de pontos para base de dados 670/2014.
julia> scatter(validacao.d, validacao.b, facecolor= "none", edgecolors="blue",
s=20, label="Base 98/2012") #nuvem de pontos para base de dados 98/2012.
julia> xlabel("Diâmetro à altura do peito (cm)")
julia> ylabel("Biomassa (Kg)")
julia> legend(prop=Dict("size"=>10))
julia> # Gráfico de Altura
julia> scatter(dados.h, dados.b, facecolor= "none", edgecolors="orange", s=20,
label="Base 670/2014")
julia> scatter(validacao.h, validacao.b, facecolor= "none", edgecolors="blue",
s=20, label="Base 98/2012")
julia> xlabel("Altura (m)")
julia> ylabel("Biomassa (kg)")
julia> legend(prop=Dict("size"=>10))
```

FONTE: O autor (2020).

FIGURA 19 –RELAÇÃO ALOMÉTRICA DA BIOMASSA DE ÁRVORES DE ACÁCIA-NEGRA COM AS VARIÁVEIS  $d$  (cm) e  $h$  (m) PARA A BASE DE DADOS 670/2014 E PARA A98/2012.



FONTE: O autor (2020).

Observação: Ressalta-se que a base de dados 670/2014 foi utilizada para o treinamento dos algoritmos e também dos modelos de regressão e a base de dados 98/2012 para a validação.

Foi evidenciado que a base de dados de validação sobrepõe o domínio das variáveis independentes diâmetro à altura do peito e altura do conjunto de dados que foi utilizado para o treinamento, o que é desejado, ainda, que as relações não lineares da biomassa com as variáveis independentes foram observadas para as duas bases. Ressalta-se que a relação da biomassa com a variável diâmetro à altura do peito e altura possui potência maior para o conjunto de validação (Figura 20), o que vai implicar na avaliação da generalidade dos modelos testados, conforme será discutido na sequência do trabalho.

Ressalta-se que na Figura 19 a relação alométrica da biomassa com a variável diâmetro à altura do peito e altura da base de dados de validação está no domínio da base de treinamento. Entretanto, devido às características do povoamento e do ambiente afetarem a relação, a relação alométrica da base de validação possui uma potência no modelo ( $\beta_2$ ) de *power-law* ( $b = \beta_1 d^{\beta_2}$ ) maior, sobressaindo-se do valor médio da base de dados treinamento, o que nota-se na Figura 20. Assume-se que modelos mais genéricos conseguirão captar essa variabilidade, logo, terão melhor desempenho na base de dados de validação. O Algoritmo 7 descreve a criação dos gráficos utilizando-se o *package Gadfly*.

ALGORITMO7 – PROGRAMAÇÃO PARA CRIAÇÃO DOS GRÁFICOS COM LINHA DE TENDÊNCIA DAS BASE DE DADOS **670/2014** E **98/2012** PARA COMPARAÇÃO NA LINGUAGEM DE PROGRAMAÇÃO **JULIA**.

```
julia> # Gráfico
```

```

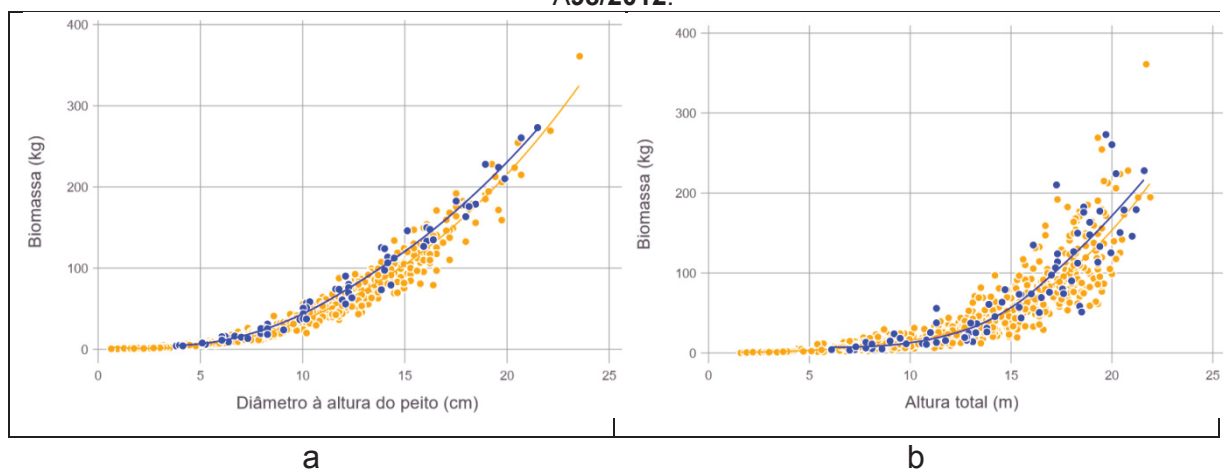
julia> plot(layer(x=validacao.d, y=validacao.b,
Geom.smooth(method=:loess,smoothing=0.9), Geom.point,
Theme(default_color="blue")), layer(x=dados.d, y=dados.b,
Geom.smooth(method=:loess, smoothing=0.9), Geom.point,
Theme(default_color="orange")), Guide.xlabel("Diâmetro à altura do peito
(cm)"), Guide.ylabel("Biomassa (kg)")) # criação do gráfico de linha de
tendência para a variável diâmetro à altura do peito. A expressão layer
define que tudo que estiver dentro dos () será considerado uma das camadas
do gráfico. Geom.smooth cria a linha de suavização (ou linha de tendência).
Geom.point gera a nuvem de pontos. Default_color define a cor.

julia> plot(layer(x=validacao.h, y=validacao.b,
Geom.smooth(method=:loess,smoothing=0.9), Geom.point,
Theme(default_color="blue")), layer(x=dados.h, y=dados.b,
Geom.smooth(method=:loess, smoothing=0.9), Geom.point,
Theme(default_color="orange")), Guide.xlabel("Altura total (m)"),
Guide.ylabel("Biomassa (kg)")) # criação do gráfico para a variável altura
total

```

FONTE: O autor (2020).

FIGURA 20 – LINHAS DE TENDÊNCIA DA DISPERSÃO DA BIOMASSA DE ÁRVORES DE ACÁCIA-NEGRA COM AS VARIÁVEIS  $d$  (cm) e  $h$  (m) PARA A BASE DE DADOS 670/2014 E PARA A98/2012.



FONTE: O autor (2020).

LEGENDA: Azul é a base de dados 98/2012; Laranja é a base de dados 670/2014.

#### 4.2 ESTATÍSTICAS VIÉS, PRECISÃO E ACURÁCIA E GRÁFICO DE DISPERSÃO

Para obtenção das estatísticas viés, precisão e acurácia, o algoritmo foi desenvolvido utilizando-se os *packages: Distributions* e *Statistics*. Nesse procedimento, utilizou-se um arquivo *.txt* intitulado estimadores, no qual as primeiras colunas são as variáveis:  $d$ ,  $h$ ,  $i$  e  $b$  e na sequência os valores de biomassa estimado

do primeiro ao  $n$ -ésimo estimador, Figura 21. Uma ressalva é que na Figura 22 os dados estão ordenados dos menores valores de biomassa para os maiores, esse procedimento foi realizado puramente para observar a evolução dos resultados, não acarretando em nenhuma mudança nas análises.

FIGURA 21 – BASE DE DADOS 670/2014 CONTENDO ESTIMADORES DE BIOMASSA POR ESTIMADORES GENÉRICOS E SEUS RESPECTIVOS RESULTADOS PARA CADA ÁRVORE DE ACÁCIA-NEGRA PARA EXEMPLIFICAR.

Row	d Float64	h Float64	i Float64	b Float64	m23 Float64	m35 Float64	m37 Float64	m39 Float64
1	0.64	1.9	1.0	0.75	0.0354311	0.810213	2.85817	0.403985
2	0.64	1.6	1.0	0.31	0.0324655	0.428843	2.85184	0.407083
3	0.64	1.65	1.0	0.42	0.0326809	0.49486	2.85293	0.406486
4	0.95	1.8	1.0	0.78	0.0482492	1.31228	2.86447	0.404543
5	0.95	1.9	1.0	0.61	0.0514094	1.40227	2.86632	0.404039
6	1.27	2.2	1.0	1.25	0.0946766	1.95087	2.87922	0.407184
7	1.27	2.35	1.0	0.77	0.105829	2.00696	2.88163	0.407674
8	1.27	2.2	1.0	0.67	0.0946766	1.95087	2.87922	0.407184
9	1.27	2.3	1.0	0.71	0.10197	1.98918	2.88083	0.407488
10	1.27	2.4	1.0	0.77	0.10983	2.02388	2.88243	0.407882
11	1.43	2.5	1.0	0.95	0.141111	2.13218	2.88785	0.412432
12	1.43	2.3	1.0	1.03	0.122326	2.08297	2.88465	0.411068
13	1.59	2.6	1.0	1.17	0.178666	2.20582	2.89351	0.4188
14	1.59	2.55	1.0	1.29	0.172719	2.19812	2.89268	0.41823
15	1.59	2.5	1.0	1.39	0.166942	2.19004	2.89185	0.417687
16	1.59	2.4	1.0	0.78	0.155891	2.17262	2.89021	0.416675
17	1.75	2.8	1.0	1.15	0.237904	2.26726	2.90142	0.429088
18	1.75	2.7	1.0	1.47	0.223183	2.25719	2.89961	0.427399
653	17.98	16.4	5.25	132.73	143.182	149.401	151.229	155.498
654	17.98	18.25	10.08	165.14	159.461	169.09	159.617	162.792
655	18.14	20.6	5.75	173.25	143.213	151.455	150.239	148.664
656	18.3	19.5	10.75	179.41	160.576	170.649	160.862	162.805
657	18.46	19.5	5.75	155.99	150.173	157.904	155.87	157.755
658	18.94	18.8	9.83	185.17	166.138	174.092	164.043	173.1
659	18.94	19.3	10.08	190.44	165.792	173.957	163.932	172.048
660	19.1	21.3	10.08	194.6	164.325	172.65	163.703	169.85
661	19.1	21.9	10.75	194.87	163.718	172.523	163.581	167.592
662	19.26	20.8	10.08	228.12	166.009	174.07	164.464	172.462
663	19.42	19.8	10.08	212.72	167.717	175.51	165.182	175.263
664	19.58	19.7	9.83	171.7	168.291	175.896	165.531	176.582
665	19.74	16.7	10.08	159.2	169.621	177.082	165.939	179.169
666	19.74	20.2	10.75	206.09	168.774	176.643	165.805	176.133
667	20.37	20.4	9.83	223.79	170.095	177.477	166.676	179.652
668	20.53	19.5	10.75	254.51	170.748	178.235	166.936	180.247
669	20.69	19.6	10.08	215.04	170.871	178.203	167.069	180.995
670	22.12	19.3	9.83	269.25	171.874	179.117	167.753	183.011
671	23.55	21.7	9.83	361.07	172.184	179.473	167.986	183.478

FONTE: O autor (2020).

LEGENDA:  $i$  = idade das árvores;  $d$  = diâmetro à altura do peito;  $h$  = altura total;  $b$  = biomassa total;  $m23, m35, m37$  e  $m39$  = biomassa estimada por estimadores genéricos.

#### ALGORITMO8 – REPRESENTAÇÃO DA PROGRAMAÇÃO DAS ESTATÍSTICAS VIÉS, PRECISÃO E ACURÁCIA, DENOMINADA FUNÇÃO VPA

```
julia> # Definir parâmetros
julia> Y = estimadores.b; # ao se utilizar a expressão ";" define-se que o
                        resultado da ação será omitido.
julia> Ye = estimadores.[:,5:end]# definição de uma matriz de n colunas com
                        todos os estimadores.
julia> p = size(Ye,2)# número de colunas
julia> #Função VPA
julia> function VPA(dados)
    ei = [Y-Ye[:,i] for i in 1:p] # resíduo; nesse procedimento utiliza-se
                        uma função de loop,que consiste em criar uma função de causa e
```



consequências traduzido como (para  $i$  dentro de 1 até  $p$ ). Nesse processamento, os resíduos são calculados para cada coluna, ou seja, para cada modelo treinado.

`ebar = [(sum(Y-Ye[:,i])/length(Y)) for i in 1:p]` # cálculo do viés acumulado; *sum* refere-se a somatória; *length* refere-se a dimensão de uma coluna.

`ebarp = [(((sum(Y-Ye[:,i])/length(Y))/mean(Y))*100) for i in 1:p]` # cálculo do viés relativo.

`Se=[sqrt((sum(((Y-Yhat[:,i]).-((sum(Y-Ye[:,i])/length(Y))))).^2))/(length(Y)-1)) for i in 1:p]` # cálculo da precisão acumulada; o ponto após uma operação matemática indica que essa operação será realizada para cada repetição (para cada árvore); *sqrt* é a operação matemática de raiz quadrada (*square root*).

`Sep=[(sqrt((sum(((Y-Yhat[:,i]).-((sum(Y-Ye[:,i])/length(Y))))).^2))/(length(Y)-1))/mean(Y)*100) for i in 1:p]`

`mx=[sqrt(((sqrt((sum(((Y-Yhat[:,i]).-((sum(Y-Yhat[:,i])/length(Y))))).^2))/(length(Y)-1))^2+(((sum(Y-Yhat[:,i])/length(Y))^2))) for i in 1:p]`

`mxp=[(((sqrt(((sqrt((sum(((Y-Yhat[:,i]).-((sum(Y-Yhat[:,i])/length(Y))))).^2))/(length(Y)-1))^2+(((sum(Y-Yhat[:,i])/length(Y))^2)))/(mean(Y))*100) for i in 1:p]`

`VPA = DataFrame(Estimador=names(m[:,5:end]), Viés=ebar, ViéP=ebarp, Precisão=Se, PrecisãoP=Sep, Acurácia=mx, AcuráciaP=mxp)` # criação de um *dataframe* em que nas colunas são definidas as estatísticas e nas linhas os valores dos estimadores; a vírgula entre cada variável define-se que está se criando uma nova coluna.

`end`

`julia> #Rodar a função VPA`

`julia> VPA(estimador)`

FONTE: O autor (2020).

Conforme o Algoritmo 9, foram construídos os gráficos de dispersão, utilizando o pacote *PyPlot*.

ALGORITMO9 – REPRESENTAÇÃO DA PROGRAMAÇÃO DOS GRÁFICOS DE DISPERSÃO PARA OS ESTIMADORES DE BIOMASSA EM LINGUAGEM DE PROGRAMAÇÃO *JULIA*.

`julia> # Gráfico de dispersão`

`julia> scatter(dados.d, dados.b, facecolor= "none", edgecolors="orange", s=20, label="Biomassa total")` # *dados.d* será o eixo x, *dados.b* será o eixo y, *facecolor* com a expressão *none* denota para que os pontos do gráfico sejam



```

vazados, edgecolor é a cor da borda dos pontos, nesse caso laranja; s é a
dimensão dos pontos.
julia> scatter(dados.d, RF1, facecolor= "none", edgecolors="blue", s=20,
label="Floresta Aleatória 1")
julia> xlabel("Diâmetro à altura do peito (cm)")# número de colunas
julia> ylabel("Biomassa (kg)")
julia> title("Floresta Aleatória 1")
julia> legend(prop=Dict("size"=>10))# a expressão após Legend descrevem como
dimensionar a legenda.

```

FONTE: O autor (2020).

No Algoritmo 10 tem-se a programação para criação dos gráficos de resíduos.

ALGORITMO10 – REPRESENTAÇÃO DA PROGRAMAÇÃO DOS GRÁFICOS DE RESÍDUOS PARA OS ESTIMADORES DE BIOMASSA EM LINGUAGEM DE PROGRAMAÇÃO *JULIA*.

```

julia> # Gráfico de resíduo
julia> residuo1 = dados.b - rede1;# criação do resíduo da primeira rede neural
(esse procedimento foi utilizado para aplicação nas outras técnicas de
aprendizagem de máquina).
julia> residuo2 = dados.b - rede2;# expressão para separa os resultados dos quatro
melhores estimadores de um algoritmo.
julia> residuo3 = dados.b - rede3;
julia> residuo4 = dados.b - rede4;# ressalta-se que o processo de plotagem foi
adaptado para cada um dos resíduos;
julia> plot(X.d, residuo1, "b.") # dados.d será o eixo x; residuo1 será o eixo
y; b. é a cor, azul, neste caso.
julia> xlabel("Diâmetro à altura do peito (cm)")
ylabel("Resíduo (Kg)")
title("Rede Neural Artificial (momentum = 0.6, learning_rate_init =
0.03)")
ylim(-100, 100) # dimensionamento do eixo y
xlim(0, 25) # dimensionamento do eixo x
plt.plot([0, 25],[0,0],'k') # criação da linha no eixo x

```

FONTE: O autor (2020).

## 4.3 TÉCNICAS DE APRENDIZAGEM DE MÁQUINA USANDO LINGUAGEM DE PROGRAMAÇÃO *JULIA*

### 4.3.1 Redes neurais artificiais

Para utilizar a técnica de redes neurais artificiais utilizaram-se os seguintes *packages*: *DataFrames*, *ScikitLearn*, *Statistics*, *Distributions* e *XLSX*. No Algoritmo 12 é apresentada a programação de redes neurais artificiais, no qual a descrição dos termos pode ser encontrada em Scikitlearn (2020a).

ALGORITMO 11 – REPRESENTAÇÃO DA PROGRAMAÇÃO DE REDES NEURAIS ARTIFICIAIS PARA ESTIMAR A BIOMASSA EM LINGUAGEM DE PROGRAMAÇÃO JULIA.

```
julia> # Estimadores
julia> @sk_import neural_network: MLPRegressor # Importar o multilayer perceptron regressor da biblioteca ScikitLearn.
julia> modelo_padrao = MLPRegressor()# expressão para gerar um estimador padrão do algoritmo.
julia> MLPRegressor(hidden_layer_sizes=(100,),activation='relu',solver='adam',alpha=0.0001,batch_size='auto',learning_rate='constant',learning_rate_init=0.001,power_t=0.5,max_iter=200,shuffle=True,random_state=None,tol=0.0001,verbose=False,warm_start=False,momentum=0.9,nesterovs_momentum=True,early_stopping=False,validation_fraction=0.1,beta_1=0.9,beta_2=0.999,epsilon=1e-08,n_iter_no_change=10)# informações da rede neural padrão.
julia> model1 = MLPRegressor(activation="logistic", solver="lbfgs", momentum=0.3, max_iter=1000, hidden_layer_sizes=(15, 15), learning_rate_init=0.01); # estimador 1; ao se acrescentar alguma informação dentro do algoritmo MLPRegressor está se alterando a configuração padrão; hidden_layer_size demonstra que possui duas camadas ocultas com 15 neurônios cada; os parâmetros não mencionados permanecem na forma padrão.
julia> model2 = MLPRegressor(activation="logistic", solver="lbfgs", momentum=0.4, max_iter=1000, hidden_layer_sizes=(15, 15), learning_rate_init=0.01);
julia> model3 = MLPRegressor(activation="logistic", solver="lbfgs", momentum=0.5, max_iter=1000, hidden_layer_sizes=(15, 15), learning_rate_init=0.01);
julia> model4 = MLPRegressor(activation="logistic", solver="lbfgs", momentum=0.6, max_iter=1000, hidden_layer_sizes=(15, 15), learning_rate_init=0.01);
julia> model5 = MLPRegressor(activation="logistic", solver="lbfgs", momentum=0.3, max_iter=1000, hidden_layer_sizes=(15, 15), learning_rate_init=0.02);
julia> model6 = MLPRegressor(activation="logistic", solver="lbfgs", momentum=0.4, max_iter=1000, hidden_layer_sizes=(15, 15), learning_rate_init=0.02);
```

```

julia> model7 = MLPRegressor(activation="logistic", solver="lbfgs",
momentum=0.5, max_iter=1000, hidden_layer_sizes=(15, 15),
learning_rate_init=0.02);
julia> model8 = MLPRegressor(activation="logistic", solver="lbfgs",
momentum=0.6, max_iter=1000, hidden_layer_sizes=(15, 15),
learning_rate_init=0.02);
...
julia> model53 = MLPRegressor(activation="logistic", solver="lbfgs",
momentum=0.3, max_iter=1000, hidden_layer_sizes=(15, 15),
learning_rate_init=0.30);
julia> model54 = MLPRegressor(activation="logistic", solver="lbfgs",
momentum=0.4, max_iter=1000, hidden_layer_sizes=(15, 15),
learning_rate_init=0.30);
julia> model55 = MLPRegressor(activation="logistic", solver="lbfgs",
momentum=0.5, max_iter=1000, hidden_layer_sizes=(15, 15),
learning_rate_init=0.30);
julia> model56 = MLPRegressor(activation="logistic", solver="lbfgs",
momentum=0.6, max_iter=1000, hidden_layer_sizes=(15, 15),
learning_rate_init=0.30);
julia> #Ajuste dos regressores
julia> x1x2x3 = [dados.d dados.h dados.i];# criação de uma matrix (Array) com as
variáveis independentes (x);
julia> y = dados.b; # criação de uma matrix para a variável dependente (y);
julia> fit!(model1, x1x2x3, y); # ajuste da rede neural 1 para as variáveis x e
y;
julia> fit!(model2, x1x2x3, y); # o modo de ajustar um estimador é igual para
todas as técnicas de aprendizagem de máquina (sendo seguidos da seguinte
lógica de computação "fit!(estimador_de_qualquer_aprendizado_de_máquina,
variavel_x, variavel_y) "
julia> fit!(model3, x1x2x3, y);
...
julia> fit!(model55, x1x2x3, y);
julia> fit!(model56, x1x2x3, y);
julia> #Predição
julia> yest1 = predict(model1, hcat(x1x2x3)) # predição do valor de y para as
variáveis x com a rede neural 1.
julia> yest2 = predict(model2, hcat(x1x2x3)) # a predição também possui a mesma
ideia que o ajuste, sendo reproduzível para todas as técnicas de
aprendizagem de máquina; sua lógica de computação é:

```

```

    "predict(estimador_de_qualquer_aprendizado_de_máquina,
    hcat(variavel_x)".
julia> yest3 = predict(model3, hcat(x1x2x3))
...
julia> yest56 = predict(model56, hcat(x1x2x3))
julia> # Transformar em tabela conforme exemplo especificado na Figura 13
julia> m = DataFrame(d=X.d, h=X.h, i=X.idade, b=X.bt, m1=yest1, m2=yest2,
m3=yest3, m4=yest4, m5=yest5, m6=yest6, m7=yest7, m8=yest8, m9=yest9,
m10=yest10, m11=yest11, m12=yest12, m13=yest13, m14=yest14, m15=yest15,
m16=yest16, m17=yest17, m18=yest18, m19=yest19, m20=yest20, m21=yest21,
m22=yest22, m23=yest23, m24=yest24, m25=yest25, m26=yest26, m27=yest27,
m28=yest28, m29=yest29, m30=yest30, m31=yest31, m32=yest32, m33=yest33,
m34=yest34, m35=yest35, m36=yest36, m37=yest37, m38=yest38, m39=yest39,
m40=yest40, m41=yest41, m42=yest42, m43=yest43, m44=yest44, m45=yest45,
m46=yest46, m47=yest47, m48=yest48, m49=yest49, m50=yest50, m51=yest51,
m52=yest52, m53=yest53, m54=yest54, m55=yest55, m56=yest56)
julia> # Exportar para o Excel
julia> XLSX.openxlsx("C:/Pasta1/Pasta2/PastaN/nome_do_arquivo.xlsx", mode="w")
do xf# criação de um arquivo em formato xlsx, com o endereço já definido.
sheet = xf[1] # definição da planilha onde serão salvos os resultados.
XLSX.rename!(sheet, "Modelos") # declaração do nome da planilha
sheet["A1:BH1"] = ["d" "h" "i" "b" "m1" "m2" "m3" "m4" "m5" "m6" "m7" "m8"
"m9" "m10" "m11" "m12" "m13" "m14" "m15" "m16" "m17" "m18" "m19" "m20"
"m21" "m22" "m23" "m24" "m25" "m26" "m27" "m28" "m29" "m30" "m31" "m32"
"m33" "m34" "m35" "m36" "m37" "m38" "m39" "m40" "m41" "m42" "m43" "m44"
"m45" "m46" "m47" "m48" "m49" "m50" "m51" "m52" "m53" "m54" "m55" "m56"]#
criação do cabeçalho
sheet["A2:BH672"] = [m.d m.h m.i m.b m.m1 m.m2 m.m3 m.m4 m.m5 m.m6 m.m7
m.m8 m.m9 m.m10 m.m11 m.m12 m.m13 m.m14 m.m15 m.m16 m.m17 m.m18 m.m19
m.m20 m.m21 m.m22 m.m23 m.m24 m.m25 m.m26 m.m27 m.m28 m.m29 m.m30 m.m31
m.m32 m.m33 m.m34 m.m35 m.m36 m.m37 m.m38 m.m39 m.m40 m.m41 m.m42 m.m43
m.m44 m.m45 m.m46 m.m47 m.m48 m.m49 m.m50 m.m51 m.m52 m.m53 m.m54 m.m55
m.m56] # obtenção dos resultados dos estimadores
end

```

FONTE: O autor (2020).

Os resultados do viés, precisão e acurácia para as redes neurais artificiais estão descritos na Tabela 7. Vale ressaltar que estão descritos somente as redes neurais artificiais que apresentaram resultados de biomassa com valores positivos,

visto que valores da variável biomassa não ocorrem no domínio negativo. A sequência de apresentação dos resultados está em função dos melhores valores de viés, precisão e acurácia.

TABELA 7 – RESULTADOS DE VIÉS, PRECISÃO E ACURÁCIA PARA OS ESTIMADORES DA VARIÁVEL BIOMASSA DE ÁRVORES DE ACÁCIA-NEGRA MODELOS DE REDES NEURAIS ARTIFICIAIS, UTILIZANDO-SE A BASE DE DADOS PARATREINAMENTO670/2014.

Estimador	$\bar{e}$	$\bar{e}_{\%}$	$S_e$	$S_{e\%}$	$m_x$	$m_{x\%}$
Redes Neurais Artificiais (momentum=0.6, learning_rate_init=0.03)	-0,1790	-0,39	7,009	15,40	7,011	15,40
Redes Neurais Artificiais (momentum=0.6, learning_rate_init=0.30)	-0,0650	-0,14	7,208	15,83	7,209	15,84
Redes Neurais Artificiais (momentum=0.4, learning_rate_init=0.09)	-0,0494	-0,11	7,441	16,35	7,441	16,35
Redes Neurais Artificiais (momentum=0.3, learning_rate_init=0.25)	0,1805	0,40	7,491	16,46	7,493	16,46
Redes Neurais Artificiais (momentum=0.3, learning_rate_init=0.15)	-0,0052	-0,01	7,508	16,49	7,508	16,49
Redes Neurais Artificiais (momentum=0.4, learning_rate_init=0.08)	-0,1028	-0,23	7,523	16,53	7,524	16,53
Redes Neurais Artificiais (momentum=0.5, learning_rate_init=0.10)	-0,0856	-0,19	7,545	16,58	7,546	16,58
Redes Neurais Artificiais (momentum=0.4, learning_rate_init=0.10)	-0,0777	-0,17	7,624	16,75	7,625	16,75
Redes Neurais Artificiais (momentum=0.3, learning_rate_init=0.06)	-0,0084	-0,02	7,757	17,04	7,757	17,04
Redes Neurais Artificiais (momentum=0.3, learning_rate_init=0.01)	-0,0068	-0,01	7,811	17,16	7,811	17,16
Redes Neurais Artificiais (momentum=0.4, learning_rate_init=0.07)	0,2045	0,45	7,818	17,17	7,821	17,18
Redes Neurais Artificiais (momentum=0.6, learning_rate_init=0.10)	-0,1472	-0,32	7,829	17,20	7,830	17,20
Redes Neurais Artificiais (momentum=0.6, learning_rate_init=0.01)	-0,4669	-1,03	7,843	17,23	7,857	17,26
Redes Neurais Artificiais (momentum=0.3, learning_rate_init=0.04)	0,0145	0,03	7,991	17,55	7,991	17,55
Redes Neurais Artificiais (momentum=0.5, learning_rate_init=0.09)	-0,2970	-0,65	7,996	17,57	8,002	17,58
Redes Neurais Artificiais (momentum=0.5, learning_rate_init=0.07)	0,0667	0,15	8,010	17,60	8,010	17,60
Redes Neurais Artificiais (momentum=0.3, learning_rate_init=0.08)	0,2798	0,61	8,027	17,63	8,032	17,65
Redes Neurais Artificiais (momentum=0.3, learning_rate_init=0.09)	0,0101	0,02	8,054	17,69	8,054	17,69
Redes Neurais Artificiais (momentum=0.4, learning_rate_init=0.02)	0,2150	0,47	8,226	18,07	8,229	18,08
Redes Neurais Artificiais (momentum=0.5, learning_rate_init=0.08)	0,1874	0,41	8,234	18,09	8,236	18,09

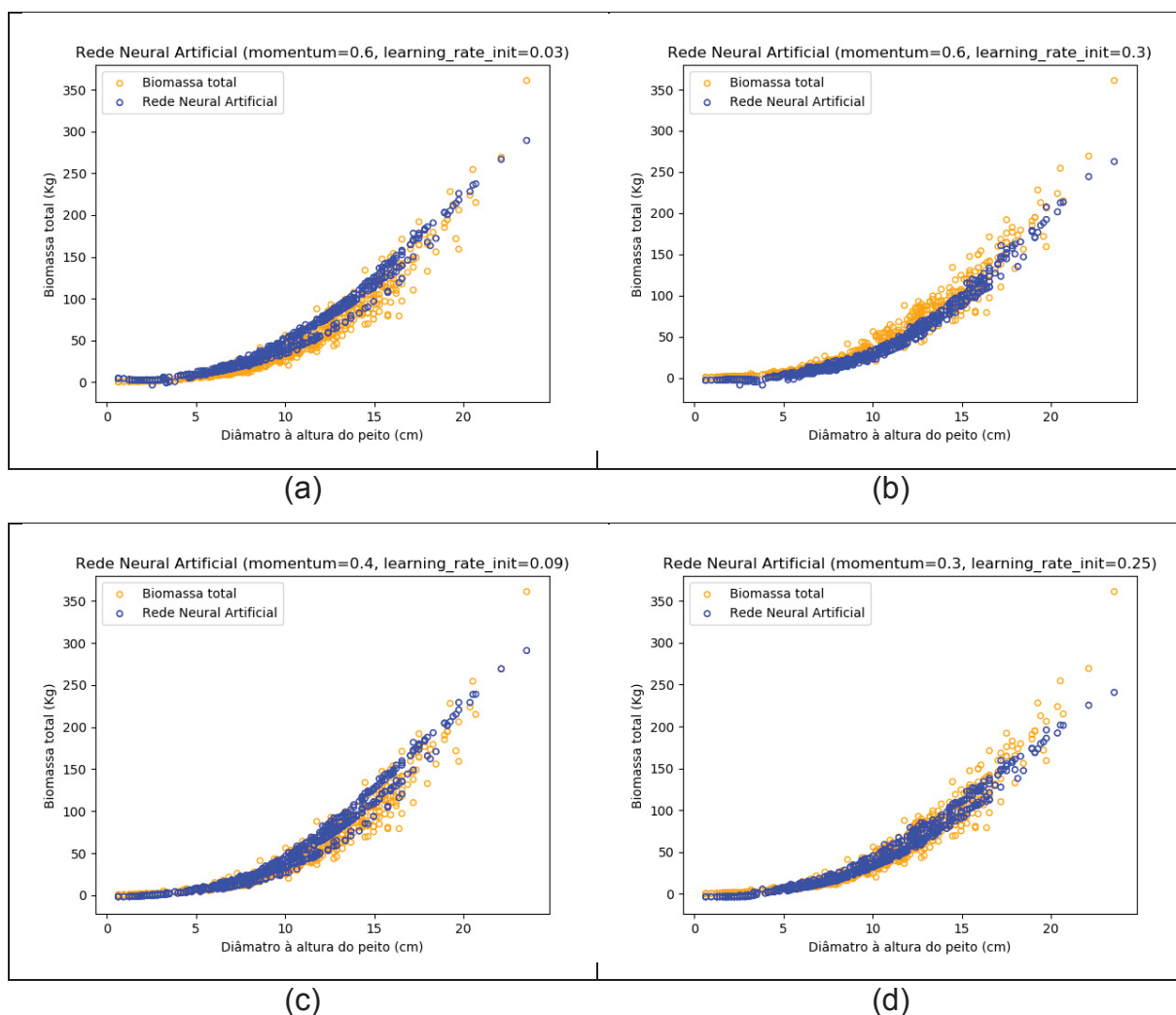
FONTE: O autor (2020).

LEGENDA:  $\bar{e}$  é o viés absoluto,  $\bar{e}_{\%}$  é o viés relativo,  $S_e$  é a precisão absoluta,  $S_{e\%}$  é a precisão relativa,  $m_x$  é a acurácia absoluta e  $m_{x\%}$  é a acurácia relativa.

Observa-se que vários dos estimadores, mas principalmente os 3 primeiras geraram viés negativo com menos de 1%, ou seja, uma subestimativa dessa magnitude. No que tange à precisão e acurácia, os quatro primeiros resultados apresentaram valores inferiores a 17%.

De modo a auxiliar a visualização dos estimadores de redes neurais artificiais, utilizou-se um modelo gráfico com os resultados das estatísticas de viés, precisão e acurácia (Figura 22).

FIGURA 22 - VALORES OBSERVADOS DA VARIÁVEL BIOMASSA DE ÁRVORES DE ACÁCIA-NEGRA E ESTIMADOS POR REDES NEURAIS ARTIFICIAIS.

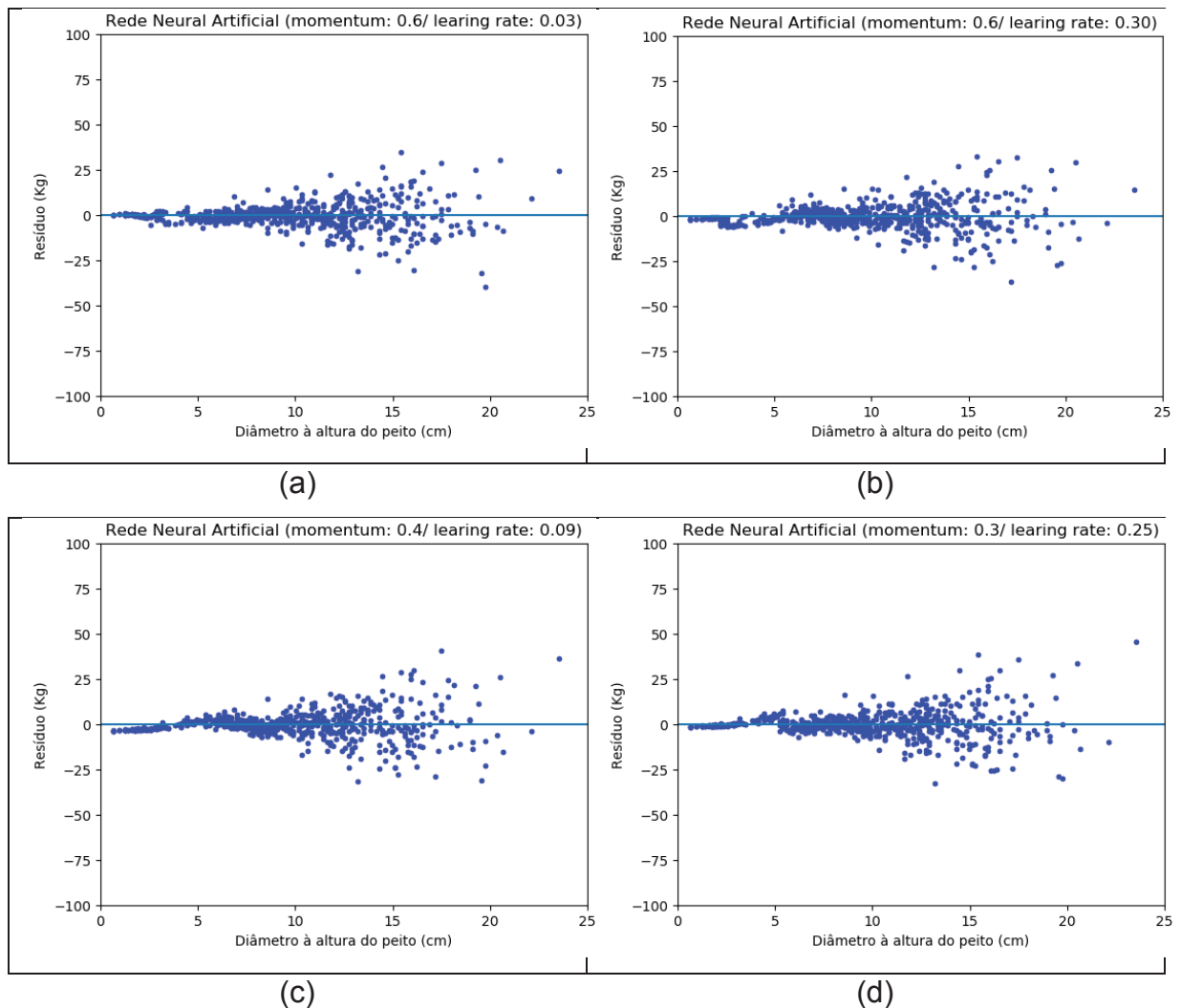


FONTE: O autor (2020).

Conforme os resultados evidenciados na Figura 22, os valores estimados são condizentes, sugerindo que os modelos de redes neurais artificiais conseguem estimar os dados de biomassa total. Nota-se também que ao se aproximar dos maiores valores da variável diâmetro à altura do peito, a qualidade na predição é menor, em termos de unidade absoluta da biomassa, contudo deve-se notar que é menor o número de dados com diâmetro à altura do peito maiores.

Os resíduos (em Kg) para os 4 estimadores com melhores resultados gerados pelas redes neurais artificiais podem ser visualizados na Figura 23. Foi observado que os valores residuais são maiores para as árvores de maior tamanho, ou seja, as árvores que possuem o valor da variável diâmetro à altura do peito entre 15 e 25 cm.

FIGURA 23– GRÁFICO DE RESÍDUOS (em kg) DA VARIÁVEL BIOMASSA DE ÁRVORES DE ACÁCIA-NEGRA MODELADA POR REDES NEURAS ARTIFICIAIS.



FONTE: O autor (2020).

#### 4.3.1.1 Predição da base de dados para validação 98/2012

Por meio do Algoritmo 12 foi realizada a predição de valores da base de dados 98/2012. Os estimadores utilizados foram os ajustados no Algoritmo 10, utilizando a base de dados 670/2014. Outra informação a ser considerada é que o procedimento será o mesmo para as outras técnicas de aprendizagem de máquina e que a base de dados de validação foi denominada de *validacao*.

ALGORITMO12 – REPRESENTAÇÃO DA PROGRAMAÇÃO DE TÉCNICAS DE APRENDIZAGEM DE MÁQUINA PARA ESTIMAR BIOMASSA DE DADOS INDEPENDENTES AOS DADOS DE AJUSTE EM LINGUAGEM DE PROGRAMAÇÃO *JULIA*.

```
julia> # Definir parâmetros
julia> x_validacao = [validacao.d validacao.h validacao.idade] # criação de uma
matriz para os valores de x.
julia> # Predição dos dados de validação
```

```
julia> Val1= predict(model1, hcat(x_validacao) # a predição dos valores seguiu o
        mesmo procedimento para todos os algoritmos.
julia> Val2= predict(model2, hcat(x_validacao) # a utilização de quatro valores
        de validação foi definido previamente a modelagem
julia> Val3= predict(model3, hcat(x_validacao)
julia> Val4= predict(model4, hcat(x_validacao)
```

FONTE: O autor (2020).

Os resultados de viés, precisão e acurácia calculados para a base de dados de validação estão apresentados na Tabela 8. Foi evidenciado que as estatísticas pioraram para a predição dos novos dados. O resultado inferior é esperado em virtude da base de dados ser diferente. Observa-se acurácia variando entre 21,91% a 33,25%.

TABELA 8 – RESULTADOS DE VIÉS, PRECISÃO E ACURÁCIA PARA OS ESTIMADORES DA VARIÁVEL BIOMASSA DE ÁRVORES DE ACÁCIA-NEGRA UTILIZANDO MODELOS DE REDES NEURAIS ARTIFICIAIS, PARA A BASE DE DADOS DE **VALIDAÇÃO 98/2012**.

Estimador	$\bar{e}$	$\bar{e}_{\%}$	$S_e$	$S_{e\%}$	$m_x$	$m_{x\%}$
Redes Neurais Artificiais (momentum=0.6, learning_rate_init=0.03)	10,31	13,77	12,76	17,04	16,40	21,91
Redes Neurais Artificiais (momentum=0.6, learning_rate_init=0.30)	11,64	15,56	14,47	19,33	18,58	24,81
Redes Neurais Artificiais (momentum=0.4, learning_rate_init=0.09)	11,70	15,63	12,92	17,25	17,43	23,28
Redes Neurais Artificiais (momentum=0.3, learning_rate_init=0.25)	14,69	19,63	20,09	26,84	24,89	33,25

FONTE: O autor (2020).

LEGENDA:  $\bar{e}$  é o viés absoluto,  $\bar{e}_{\%}$  é o viés relativo,  $S_e$  é a precisão absoluta,  $S_{e\%}$  é a precisão relativa,  $m_x$  é a acurácia absoluta e  $m_{x\%}$  é a acurácia relativa.

#### 4.3.2 *k*-vizinhos mais próximos

Para utilizar o algoritmo *k*-vizinhos mais próximos foram necessários os seguintes *packages*: *DataFrames*, *XLSX*, *ScikitLearn*, *Random* e *Statistics*. Os procedimentos para criar um arquivo *Dataframe* para os estimadores desse algoritmo e exportar os resultados para um arquivo *xlsx* são similares aos estabelecidos no Algoritmo 11. A programação de *k*-vizinhos mais próximos foi realizada conforme Algoritmo 13, com detalhamento dos termos em SCIKITLEARN (2020b).

ALGORITMO13 – REPRESENTAÇÃO DA PROGRAMAÇÃO DE *k*-VIZINHOS MAIS PRÓXIMOS PARA ESTIMAR A BIOMASSA EM LINGUAGEM DE PROGRAMAÇÃO *JULIA*.

```
julia> # Estimadores
julia> @sk_import neighbors: KNeighborsRegressor# Importar o regressor de k-
        vizinhos mais próximos da biblioteca ScikitLearn.
```



```

julia> modelo_padrao = KNeighborsRegressor()# expressão para gerar um estimador
padrao do algoritmo.
julia> KNeighborsRegressor(n_neighbors=5,weights='uniform',algorithm='auto',lea
f_size=30,p=2,metric='minkowski',metric_params=None,n_jobs=None,**kwargs
)# informações dos k-vizinhos mais próximos padrão.
julia> k01 = KNeighborsRegressor(n_neighbors=1);
julia> k02 = KNeighborsRegressor(n_neighbors=2);
julia> k03 = KNeighborsRegressor(n_neighbors=3);
...
julia> k31 = KNeighborsRegressor(n_neighbors=1, metric="manhattan");
julia> k32 = KNeighborsRegressor(n_neighbors=2, metric="manhattan");
julia> k33 = KNeighborsRegressor(n_neighbors=3, metric="manhattan");
...
julia> k61 = KNeighborsRegressor(n_neighbors=1, metric="chebyshev");
julia> k62 = KNeighborsRegressor(n_neighbors=2, metric="chebyshev");
julia> k63 = KNeighborsRegressor(n_neighbors=3, metric="chebyshev");
...
julia> K322 = KNeighborsRegressor(n_neighbors=140, metric="chebyshev",
weights="distance");
julia> K323 = KNeighborsRegressor(n_neighbors=145, metric="chebyshev",
weights="distance"););
julia> K324 = KNeighborsRegressor(n_neighbors=150, metric="chebyshev",
weights="distance"););
julia> # Ajuste dos estimadores
julia> fit!(k01, x1x2x3, y);
julia> fit!(k02, x1x2x3, y);
julia> fit!(k03, x1x2x3, y);
...
julia> fit!(k90, x1x2x3, y);
julia> # Estimar a base de dados
julia> y01 = predict(k01, hcat(x1x2x3));
julia> y02 = predict(k02, hcat(x1x2x3));
julia> y03 = predict(k03, hcat(x1x2x3));
...
julia> y89 = predict(k89, hcat(x1x2x3));
julia> y90 = predict(k90, hcat(x1x2x3));

```

FONTE: O autor (2020).

Os resultados do viés, precisão e acurácia para os  $k$ -vizinhos mais próximos estão descritos na Tabela 9. Vale ressaltar que foi excluído os resultados com o valor

$k=1$  para as três métricas, visto que nesse caso os resultados tiveram uma sobreposição (*overfitting*). A organização dos resultados se deu dos melhores valores de viés, precisão e acurácia para os piores.

TABELA 9 – RESULTADO DE VIÉS, PRECISÃO E ACURÁCIA PARA OS ESTIMADORES POR ALGORITMO DE  $k$ -VIZINHOS MAIS PRÓXIMOS.

Estimador	$\bar{e}$	$\bar{e}_{\%}$	$S_e$	$S_{e\%}$	$m_x$	$m_{x\%}$
k-Vizinhos mais próximos (n_neighbors=5, weights = "distance")	-3,71E-17	-8,14E-17	0,73	1,60	0,73	1,60
k-Vizinhos mais próximos (n_neighbors=6, weights = "distance")	-3,71E-17	-8,14E-17	0,73	1,60	0,73	1,60
k-Vizinhos mais próximos (n_neighbors=7, weights = "distance")	-3,71E-17	-8,14E-17	0,73	1,60	0,73	1,60
k-Vizinhos mais próximos (n_neighbors=8, weights = "distance")	-3,71E-17	-8,14E-17	0,73	1,60	0,73	1,60
k-Vizinhos mais próximos (n_neighbors=9, weights = "distance")	-3,44E-17	-7,56E-17	0,73	1,60	0,73	1,60
k-Vizinhos mais próximos (n_neighbors=10, weights = "distance")	-3,44E-17	-7,56E-17	0,73	1,60	0,73	1,60
k-Vizinhos mais próximos (n_neighbors=11, weights = "distance")	-3,71E-17	-8,14E-17	0,73	1,60	0,73	1,60
k-Vizinhos mais próximos (n_neighbors=12, weights = "distance")	-3,71E-17	-8,14E-17	0,73	1,60	0,73	1,60
k-Vizinhos mais próximos (n_neighbors=13, weights = "distance")	-3,71E-17	-8,14E-17	0,73	1,60	0,73	1,60
k-Vizinhos mais próximos (n_neighbors=14, weights = "distance")	-3,71E-17	-8,14E-17	0,73	1,60	0,73	1,60
k-Vizinhos mais próximos (n_neighbors=15, weights = "distance")	-3,71E-17	-8,14E-17	0,73	1,60	0,73	1,60
k-Vizinhos mais próximos (n_neighbors=16, weights = "distance")	-3,71E-17	-8,14E-17	0,73	1,60	0,73	1,60
k-Vizinhos mais próximos (n_neighbors=17, weights = "distance")	-3,71E-17	-8,14E-17	0,73	1,60	0,73	1,60
k-Vizinhos mais próximos (n_neighbors=18, weights = "distance")	-3,44E-17	-7,56E-17	0,73	1,60	0,73	1,60
k-Vizinhos mais próximos (n_neighbors=19, weights = "distance")	-3,71E-17	-8,14E-17	0,73	1,60	0,73	1,60
k-Vizinhos mais próximos (n_neighbors=20, weights = "distance")	-3,71E-17	-8,14E-17	0,73	1,60	0,73	1,60
k-Vizinhos mais próximos (n_neighbors=21, weights = "distance")	-3,44E-17	-7,56E-17	0,73	1,60	0,73	1,60
k-Vizinhos mais próximos (n_neighbors=22, weights = "distance")	-3,71E-17	-8,14E-17	0,73	1,60	0,73	1,60
k-Vizinhos mais próximos (n_neighbors=23, weights = "distance")	-3,71E-17	-8,14E-17	0,73	1,60	0,73	1,60
k-Vizinhos mais próximos (n_neighbors=24, weights = "distance")	-3,71E-17	-8,14E-17	0,73	1,60	0,73	1,60

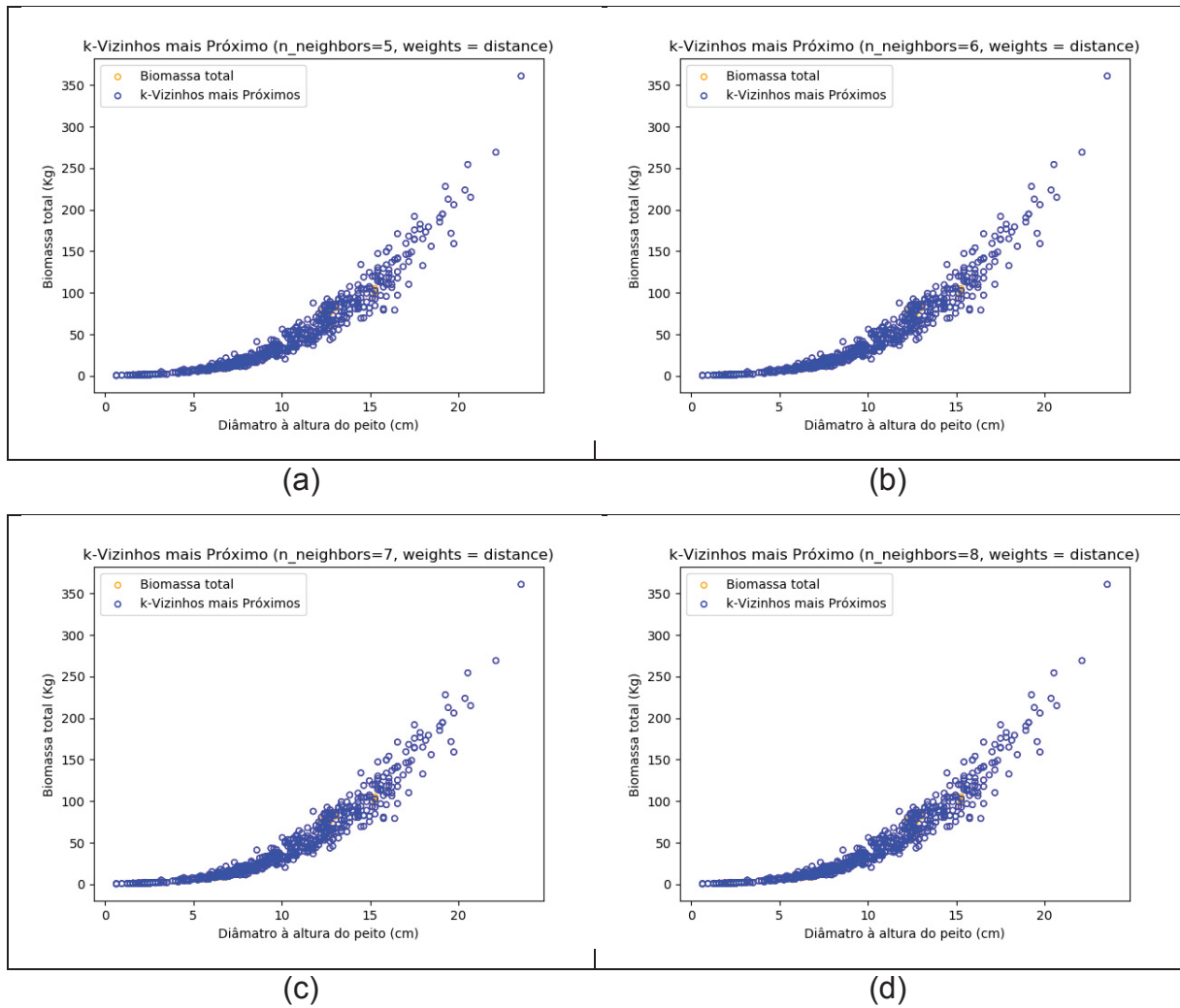
FONTE: O autor (2020).

LEGENDA:  $\bar{e}$  é o viés absoluto,  $\bar{e}_{\%}$  é o viés relativo,  $S_e$  é a precisão absoluta,  $S_{e\%}$  é a precisão relativa,  $m_x$  é a acurácia absoluta e  $m_{x\%}$  é a acurácia relativa.

Os 4 primeiros resultados de viés, apresentam valores inferiores a 1%, com resultados positivos o que acarreta em uma superestimativa de mesma intensidade. Vale ressaltar que todos os estimadores apresentaram resultados de viés positivo e a Precisão e Acurácia resultaram em valores inferiores a 16%.

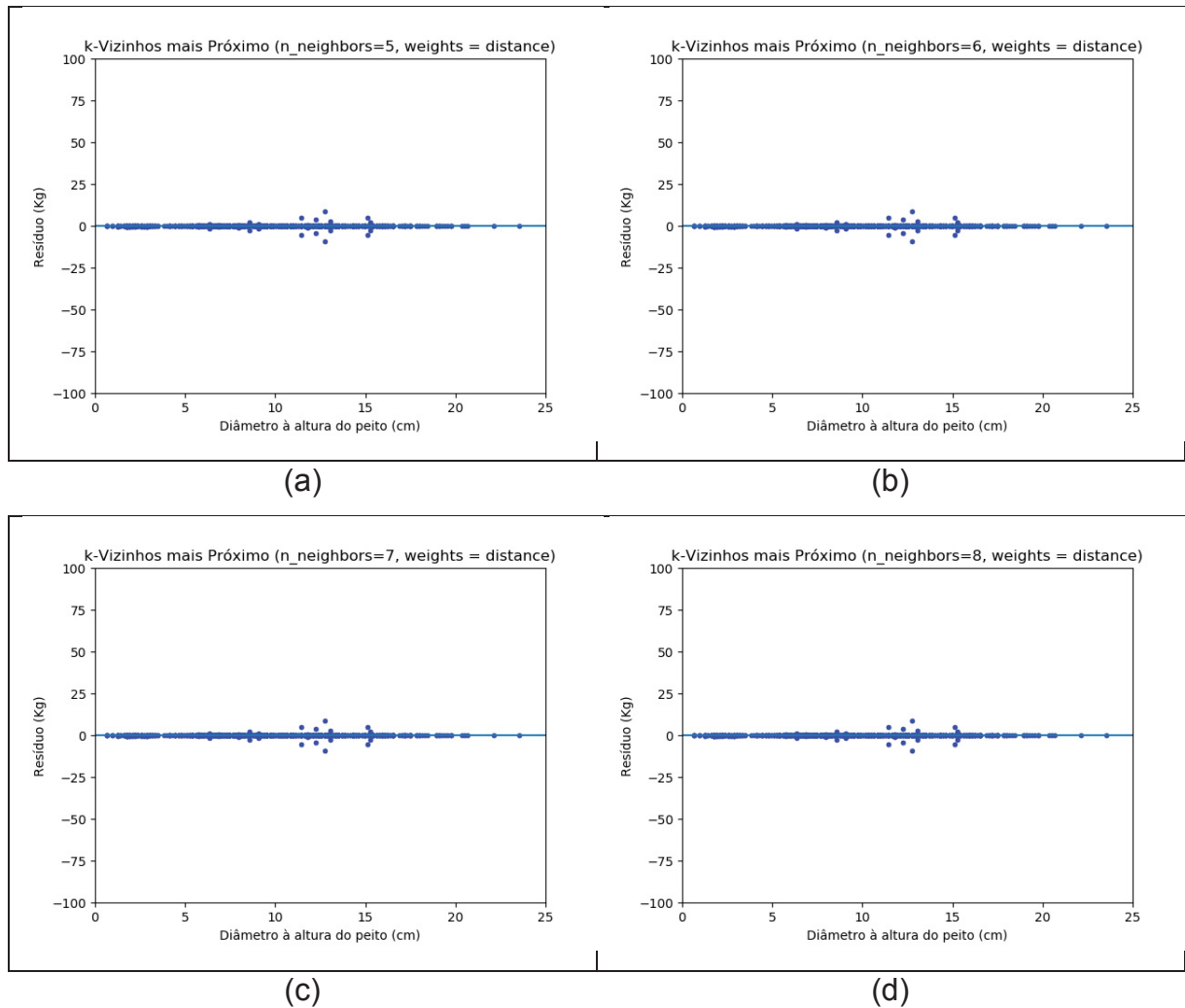
Os valores observados e preditos da biomassa, bem como os valores de resíduos (em Kg) para os 4 melhores estimadores gerados pelo algoritmo  $k$ -vizinhos mais próximos, estão apresentados nas Figuras 24 e 25 respectivamente. Conforme observou-se nas Figuras 14 e 15, destaca-se a mesma tendência de aumento do erro da estimativa para as árvores de maior altura (i.e., maior diâmetro à altura do peito), contudo, suas diferenças para os dados observados foram menores quando comparados com as redes neurais artificiais.

FIGURA 24 – VALORES OBSERVADOS DA VARIÁVEL BIOMASSA DE ÁRVORES DE ACÁCIA-NEGRA E ESTIMADOS POR  $k$ -VIZINHOS MAIS PRÓXIMOS



FONTE: O autor (2020).

FIGURA 25 – GRÁFICO DE RESÍDUOS (em kg) DA VARIÁVEL BIOMASSA DE ÁRVORES DE ACÁCIA-NEGRA MODELADA POR *k*-VIZINHOS MAIS PRÓXIMOS.



FONTE: O autor (2020).

#### 4.3.2.1 Predição da base de dados para validação 98/2012

A predição da biomassa para uma nova base de dados foi realizada da mesma forma à apresentada no Algoritmo 13. Os resultados de viés, precisão e acurácia, calculados para a base de dados de validação, estão apresentados na Tabela 10.

TABELA 10 – RESULTADOS DE VIÉS, PRECISÃO E ACURÁCIA PARA OS ESTIMADORES DA VARIÁVEL BIOMASSA DE ÁRVORES DE ACÁCIA-NEGRA UTILIZANDO MODELOS DE *k*-VIZINHOS MAIS PRÓXIMOS, PARA A BASE DE DADOS DE **VALIDAÇÃO 98/2012**.

Estimador	$\bar{e}$	$\bar{e}_{\%}$	$S_e$	$S_{e\%}$	$m_x$	$m_{x\%}$
k-Vizinhos mais próximos (n_neighbors=5, weights = "distance");	14,18	18,94	19,68	26,28	24,25	32,40
k-Vizinhos mais próximos (n_neighbors=6, weights = "distance");	14,05	18,76	18,87	25,21	23,53	31,43
k-Vizinhos mais próximos (n_neighbors=7, weights = "distance");	13,95	18,64	18,63	24,89	23,28	31,09
k-Vizinhos mais próximos (n_neighbors=8, weights = "distance");	14,34	19,16	19,15	25,58	23,93	31,96

FONTE: O autor (2020).

LEGENDA:  $\bar{e}$  é o viés absoluto,  $\bar{e}_{\%}$  é o viés relativo,  $S_e$  é a precisão absoluta,  $S_{e\%}$  é a precisão relativa,  $m_x$  é a acurácia absoluta e  $m_{x\%}$  é a acurácia relativa.

#### 4.3.3 Florestas aleatórias

Os *packages* utilizados para desenvolver a programação das florestas aleatórias são: *DataFrames*, *XLSX*, *DecisionTree*, *ScikitLearn*, *Random* e *Statistics*. No Algoritmo 14 tem-se a programação de florestas aleatórias e para melhor compreensão estão especificados os termos em Scikitlearn (2020c).

ALGORITMO14 – REPRESENTAÇÃO DA PROGRAMAÇÃO DE FLORESTAS ALEATÓRIAS PARA ESTIMAR A BIOMASSA EM LINGUAGEM DE PROGRAMAÇÃO JULIA.

```
julia> # Estimadores
julia> modelo_padrao = RandomForestRegressor()
julia> RandomForestRegressor(n_estimators='warn', criterion='mse', max_depth=None,
, min_samples_split=2, min_samples_leaf=1, min_weight_fraction_leaf=0.0, max
_features='auto', max_leaf_nodes=None, min_impurity_decrease=0.0, min_impur
ity_split=None, bootstrap=True, oob_score=False, n_jobs=None, random_state=N
one, verbose=0, warm_start=False)
julia> f01 = RandomForestRegressor(n_estimators=10);
julia> f02 = RandomForestRegressor(n_estimators=20);
julia> f03 = RandomForestRegressor(n_estimators=30);
...
julia> f20 = RandomForestRegressor(n_estimators=10, min_samples_leaf=2);
julia> f21 = RandomForestRegressor(n_estimators=20, min_samples_leaf=2);
julia> f22 = RandomForestRegressor(n_estimators=30, min_samples_leaf=2);
...
julia> f39 = RandomForestRegressor(n_estimators=10, min_samples_leaf=3);
julia> f40 = RandomForestRegressor(n_estimators=20, min_samples_leaf=3);
julia> f41 = RandomForestRegressor(n_estimators=30, min_samples_leaf=3);
...
julia> f55 = RandomForestRegressor(n_estimators=400, min_samples_leaf=3,
oob_score="True");
```

```

julia> f56 = RandomForestRegressor(n_estimators=450, min_samples_leaf=3,
oob_score="True");
julia> F109 = RandomForestRegressor(n_estimators=500, min_samples_leaf=3,
oob_score="True");
julia> # Ajuste dos estimadores
julia> fit!(f01, x1x2x3, y);
julia> fit!(f02, x1x2x3, y);
julia> fit!(f03, x1x2x3, y);
...
julia> fit!(f108, x1x2x3, y);
julia> # Estimar a base de dados
julia> y01 = predict(f01, hcat(x1x2x3));
julia> y02 = predict(f02, hcat(x1x2x3));
julia> y03 = predict(f03, hcat(x1x2x3));
...
julia> y107 = predict(f55, hcat(x1x2x3));
julia> y108 = predict(f56, hcat(x1x2x3));

```

FONTE: O autor (2020).

Os resultados do viés, precisão e acurácia para as florestas aleatórias estão descritos na Tabela 11.

TABELA 11 – RESULTADOS DE VIÉS, PRECISÃO E ACURÁCIA PARA OS ESTIMADORES DA VARIÁVEL BIOMASSA DE ÁRVORES DE ACÁCIA-NEGRA UTILIZANDO-SE MODELOS DE REDES NEURAIS ARTIFICIAIS, PARA A BASE DE DADOS DE **TREINAMENTO 670/2014**.

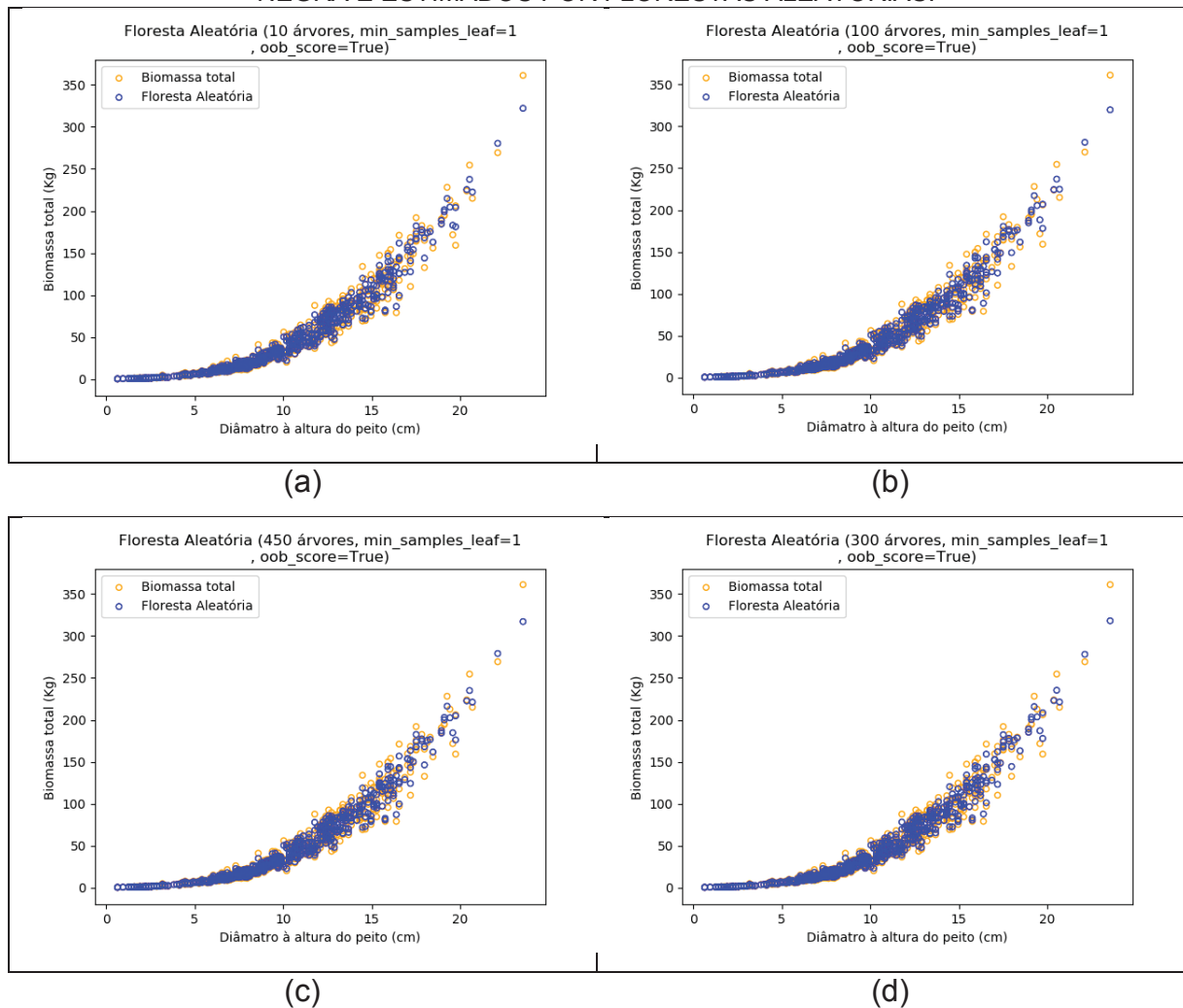
Regressor	$\bar{e}$	$\bar{e}_{\%}$	$S_e$	$S_{e\%}$	$m_x$	$m_{x\%}$
Floresta aleatória (n_estimators=60, oob_score="True")	0,03	0,06	3,62	7,96	3,62	7,96
Floresta aleatória (n_estimators=100, oob_score="True")	-0,02	-0,04	3,66	8,05	3,66	8,05
Floresta aleatória (n_estimators=450, oob_score="True")	0,01	0,02	3,69	8,11	3,69	8,11
Floresta aleatória (n_estimators=300, oob_score="True")	-0,05	-0,11	3,71	8,14	3,71	8,14
Floresta aleatória (n_estimators=200, oob_score="True")	-0,05	-0,11	3,72	8,16	3,72	8,17
Floresta aleatória (n_estimators=350)	-0,02	-0,05	3,73	8,19	3,73	8,19
Floresta aleatória (n_estimators=100)	-0,02	-0,03	3,74	8,21	3,74	8,21
Floresta aleatória (n_estimators=70)	0,00	0,00	3,74	8,21	3,74	8,21
Floresta aleatória (n_estimators=250)	-0,02	-0,05	3,75	8,23	3,75	8,23
Floresta aleatória (n_estimators=400)	-0,02	-0,05	3,76	8,26	3,76	8,26
Floresta aleatória (n_estimators=300)	-0,01	-0,01	3,76	8,26	3,76	8,26
Floresta aleatória (n_estimators=50)	0,03	0,06	3,77	8,28	3,77	8,28
Floresta aleatória (n_estimators=90)	-0,05	-0,12	3,77	8,28	3,77	8,28
Floresta aleatória (n_estimators=450)	0,02	0,04	3,78	8,30	3,78	8,30
Floresta aleatória (n_estimators=40, oob_score="True")	0,02	0,04	3,78	8,31	3,78	8,31
Floresta aleatória (n_estimators=500, oob_score="True")	-0,02	-0,04	3,80	8,34	3,80	8,34
Floresta aleatória (n_estimators=30)	0,04	0,08	3,80	8,35	3,80	8,35
Floresta aleatória (n_estimators=80, oob_score="True")	0,01	0,02	3,80	8,35	3,80	8,35
Floresta aleatória (n_estimators=150)	0,07	0,16	3,80	8,35	3,80	8,35
Floresta aleatória (n_estimators=250, oob_score="True")	0,02	0,05	3,81	8,36	3,81	8,36

FONTE: O autor (2020).

Para esse algoritmo, observou-se valores de viés próximos de zero, o que é desejado, sendo inferiores a 1%. Especificamente, os quatro primeiros resultados

foram inferiores a 0,2%. Para precisão e acurácia, todos os resultados são inferiores a 19% e os quatro primeiros inferiores a 14%. Os valores observados e preditos, para os 4 melhores estimadores gerados com as florestas aleatórias, estão apresentados na Figura 26.

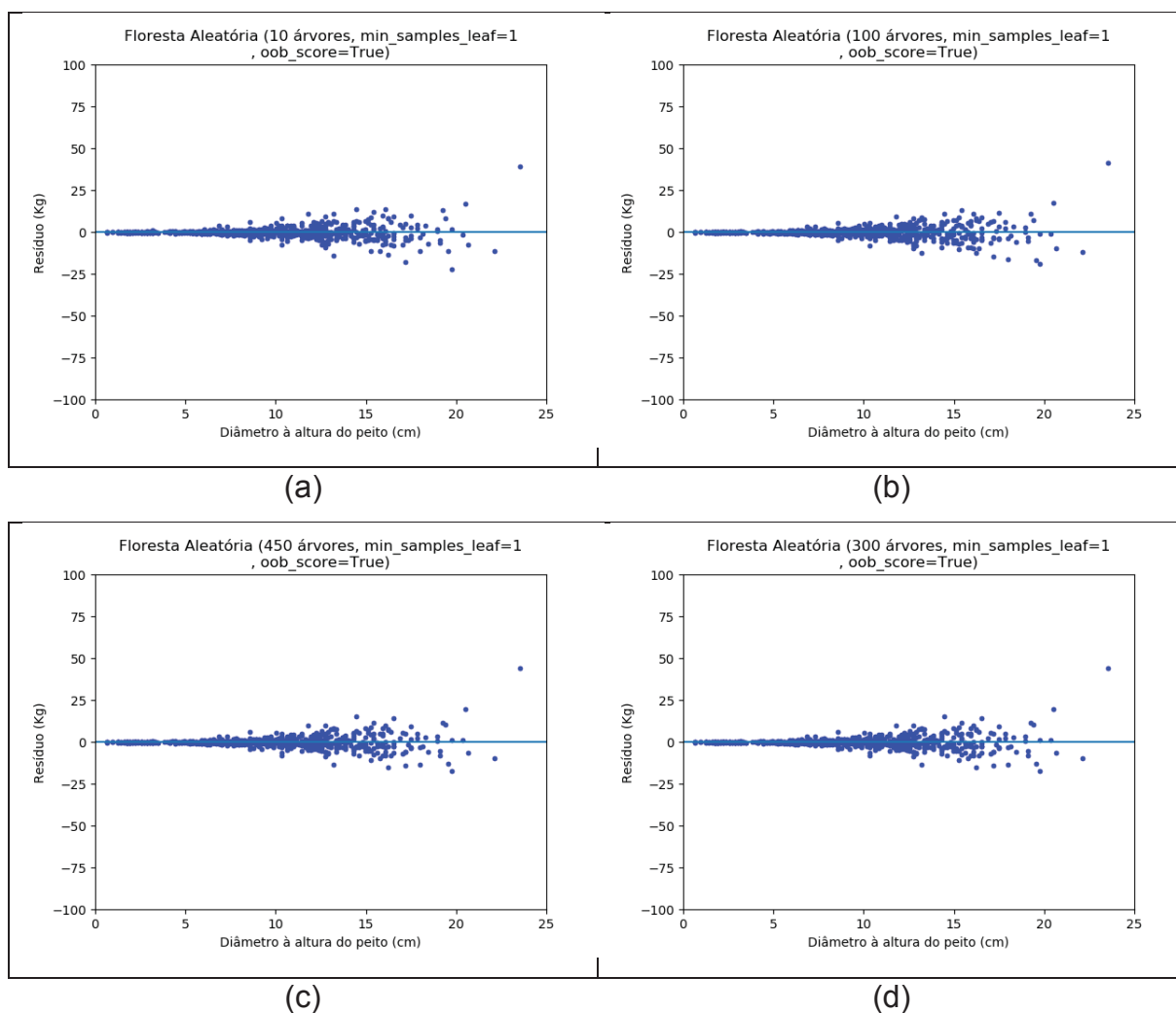
FIGURA 26– VALORES OBSERVADOS DA VARIÁVEL BIOMASSA DE ÁRVORES DE ACÁCIA-NEGRA E ESTIMADOS POR FLORESTAS ALEATÓRIAS.



FONTE: O autor (2020).

As estimativas de biomassa para as maiores árvores foram realizadas com maior erro. Isso também foi observado nas redes neurais artificiais e  $k$ -vizinhos mais próximos. Os resíduos (em Kg) com 4 estimadores usando florestais aleatórias podem ser visualizados na Figura 27.

FIGURA 27– GRÁFICO DE RESÍDUOS (em Kg) DA VARIÁVEL BIOMASSA DE ÁRVORES DE ACÁCIA-NEGRA MODELADA POR FLORESTAS ALEATÓRIAS.



FONTE: O autor (2020).

#### 4.3.3.1 Predição da base de dados para validação 98/2012

Os resultados de viés, precisão e acurácia, calculados para a base de dados de validação, estão apresentados na Tabela 12.

TABELA 12 – RESULTADOS DE VIÉS, PRECISÃO E ACURÁCIA PARA OS ESTIMADORES DA VARIÁVEL BIOMASSA DE ÁRVORES DE ACÁCIA-NEGRA UTILIZANDO MODELOS DE FLORESTAS ALEATÓRIAS, PARA A BASE DE DADOS DE **VALIDAÇÃO 98/2012**.

Estimador	$\bar{e}$	$\bar{e}_{\%}$	$S_e$	$S_{e\%}$	$m_x$	$m_{x\%}$
Floresta aleatória (n_estimators=60, oob_score="True")	10,36	13,85	11,12	14,86	15,20	20,31
Floresta aleatória (n_estimators=100, oob_score="True")	9,96	13,31	10,89	14,55	14,76	19,72
Floresta aleatória (n_estimators=450, oob_score="True")	10,23	13,67	11,32	15,12	15,26	20,38
Floresta aleatória (n_estimators=300, oob_score="True")	9,97	13,32	10,97	14,65	14,83	19,80

FONTE: O autor (2020).

#### 4.3.4 Máquinas de vetores de suporte



Os *packages* utilizados para desenvolver a programação das máquinas de vetores de suporte foram: *DataFrames*, *XLSX*, *ScikitLearn*, *Random* e *Statistics*. A modelagem da biomassa por meio dessa técnica de aprendizagem de máquina foi realizada conforme Algoritmo 15. Os parâmetros estão descritos em Scikitlearn (2020d):

ALGORITMO15 – REPRESENTAÇÃO DA PROGRAMAÇÃO DE MÁQUINAS DE VETORES DE SUPORE PARA ESTIMAR A BIOMASSA EM LINGUAGEM DE PROGRAMAÇÃO JULIA.

```
julia> # Estimadores
julia> @sk_import svm: SVR
julia> modelo_padrao = SVR()
julia> SVR(kernel='rbf', degree=3, gamma='scale', coef0=0.0, tol=0.001, C=1.0, epsilon=0.1, shrinking=True, cache_size=200, verbose=False, max_iter=-1)
julia> svm01 = SVR(gamma=2^-8 , C=2^1);
julia> svm02 = SVR(gamma=2^-8 , C=2^2);
julia> svm03 = SVR(gamma=2^-8 , C=2^3);
...
julia> svm62 = SVR(gamma=2^-1 , C=2^6);
julia> svm63 = SVR(gamma=2^-1 , C=2^7);
julia> svm64 = SVR(gamma=2^-1 , C=2^8);
julia> svm65 = SVR(epsilon=0.5, gamma=2^-8 , C=2^1);
julia> svm66 = SVR(epsilon=0.5, gamma=2^-8 , C=2^2);
julia> svm67 = SVR(epsilon=0.5, gamma=2^-8 , C=2^3);
...
julia> svm126 = SVR(epsilon=0.5, gamma=2^-1 , C=2^6);
julia> svm127 = SVR(epsilon=0.5, gamma=2^-1 , C=2^7);
julia> svm128 = SVR(epsilon=0.5, gamma=2^-1 , C=2^8);
julia> # Ajuste dos estimadores
julia> fit!(svm01, x1x2x3, y);
julia> fit!(svm02, x1x2x3, y);
julia> fit!(svm03, x1x2x3, y);
...
julia> fit!(svm128, x1x2x3, y);
julia> # Estimar a base de dados
julia> y01 = predict(svm01, hcat(x1x2x3));
julia> y02 = predict(svm02, hcat(x1x2x3));
julia> y03 = predict(svm03, hcat(x1x2x3));
...
julia> y127 = predict(svm127, hcat(x1x2x3));
julia> y128 = predict(svm128, hcat(x1x2x3));
```

FONTE: O autor (2020).

Os resultados do viés, precisão e acurácia para as máquinas de vetores de suporte estão descritos na Tabela 13.

TABELA 13 – RESULTADOS DE VIÉS, PRECISÃO E ACURÁCIA PARA OS ESTIMADORES DA VARIÁVEL BIOMASSA DE ÁRVORES DE ACÁCIA-NEGRA UTILIZANDO-SE MODELOS DE MÁQUINAS DE VETORES DE SUPORTE, PARA A BASE DE DADOS DETREINAMENTO 670/2014.

Estimador	$\bar{e}$	$\bar{e}_{\%}$	$S_e$	$S_{e\%}$	$m_x$	$m_{x\%}$
Máquinas de vetor de suporte (epsilon=0.5, gamma=2 <sup>-1</sup> , C=2 <sup>8</sup> )	0,11	0,23	5,09	11,19	5,09	11,19
Máquinas de vetor de suporte (gamma=2 <sup>-1</sup> , C=2 <sup>8</sup> )	0,17	0,38	5,13	11,28	5,14	11,28
Máquinas de vetor de suporte (epsilon=0.5, gamma=2 <sup>-2</sup> , C=2 <sup>8</sup> )	0,17	0,37	5,68	12,47	5,68	12,48
Máquinas de vetor de suporte (gamma=2 <sup>-2</sup> , C=2 <sup>8</sup> )	0,16	0,35	5,72	12,56	5,72	12,56
Máquinas de vetor de suporte (epsilon=0.5, gamma=2 <sup>-3</sup> , C=2 <sup>8</sup> )	0,16	0,34	6,18	13,57	6,18	13,58
Máquinas de vetor de suporte (gamma=2 <sup>-3</sup> , C=2 <sup>8</sup> )	0,21	0,47	6,22	13,65	6,22	13,66
Máquinas de vetor de suporte (epsilon=0.5, gamma=2 <sup>-4</sup> , C=2 <sup>8</sup> )	0,35	0,77	6,59	14,48	6,60	14,50
Máquinas de vetor de suporte (gamma=2 <sup>-4</sup> , C=2 <sup>8</sup> )	0,36	0,79	6,61	14,52	6,62	14,55
Máquinas de vetor de suporte (epsilon=0.5, gamma=2 <sup>-5</sup> , C=2 <sup>8</sup> )	0,17	0,38	7,21	15,84	7,21	15,84
Máquinas de vetor de suporte (gamma=2 <sup>-5</sup> , C=2 <sup>8</sup> )	0,2	0,43	7,26	15,94	7,26	15,95
Máquinas de vetor de suporte (epsilon=0.5, gamma=2 <sup>-6</sup> , C=2 <sup>8</sup> )	0,49	1,09	7,81	17,15	7,82	17,18
Máquinas de vetor de suporte (gamma=2 <sup>-6</sup> , C=2 <sup>8</sup> )	0,45	0,99	7,85	17,25	7,87	17,28
Máquinas de vetor de suporte (gamma=2 <sup>-7</sup> , C=2 <sup>8</sup> )	0,41	0,91	7,86	17,27	7,87	17,29
Máquinas de vetor de suporte (epsilon=0.5, gamma=2 <sup>-7</sup> , C=2 <sup>8</sup> )	0,46	1,01	7,86	17,27	7,87	17,30
Máquinas de vetor de suporte (epsilon=0.5, gamma=2 <sup>-3</sup> , C=2 <sup>7</sup> )	0,41	0,89	7,88	17,30	7,89	17,33
Máquinas de vetor de suporte (gamma=2 <sup>-3</sup> , C=2 <sup>7</sup> )	0,46	1,01	7,9	17,35	7,91	17,38
Máquinas de vetor de suporte (epsilon=0.5, gamma=2 <sup>-4</sup> , C=2 <sup>7</sup> )	0,52	1,14	8,01	17,59	8,02	17,62
Máquinas de vetor de suporte (gamma=2 <sup>-4</sup> , C=2 <sup>7</sup> )	0,57	1,24	8,04	17,65	8,06	17,7
Máquinas de vetor de suporte (epsilon=0.5, gamma=2 <sup>-8</sup> , C=2 <sup>8</sup> )	0,50	1,09	8,06	17,71	8,08	17,74
Máquinas de vetor de suporte (gamma=2 <sup>-8</sup> , C=2 <sup>8</sup> )	0,48	1,05	8,07	17,74	8,09	17,77

FONTE: O autor (2020).

Após avaliação dos valores de viés, precisão e acurácia, foram selecionados os quatro melhores aprendizados e foi realizado um refinamento do treinamento com a segunda varredura, conforme mencionado anteriormente. Os resultados do refinamento estão apresentados na Tabela 14.

TABELA 14 – RESULTADOS DE VIÉS, PRECISÃO E ACURÁCIA PARA OS ESTIMADORES DA VARIÁVEL BIOMASSA DE ÁRVORES DE ACÁCIA NEGRA UTILIZANDO-SE MODELOS DE MÁQUINAS DE VETORES DE SUPORTE APÓS UM REFINAMENTO, PARA A BASE DE DADOS DE TREINAMENTO 670/2014

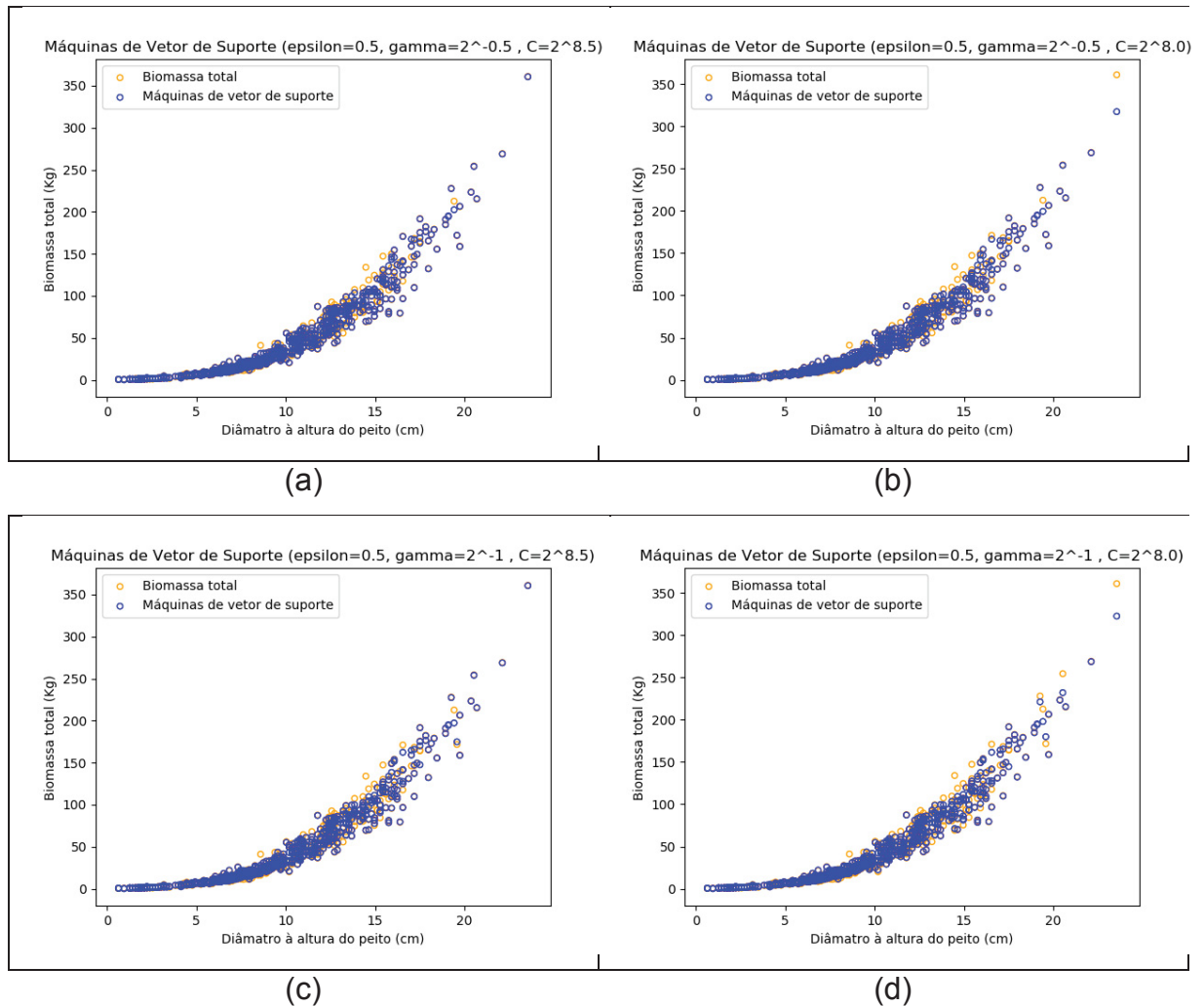
Estimador	$\bar{e}$	$\bar{e}_{\%}$	$S_e$	$S_{e\%}$	$m_x$	$m_{x\%}$
-----------	-----------	----------------	-------	-----------	-------	-----------

Máquinas de vetor de suporte ( $\gamma=2^{-0.5}$ , $C=2^{8.5}$ )	0,01	0,02	4,01	8,81	4,01	8,81
Máquinas de vetor de suporte ( $\gamma=2^{-0.5}$ , $C=2^8$ )	0,14	0,30	4,57	10,05	4,58	10,05
Máquinas de vetor de suporte ( $\gamma=2^{-1.0}$ , $C=2^{8.5}$ )	0,09	0,19	4,67	10,26	4,67	10,26
Máquinas de vetor de suporte ( $\gamma=2^{-1}$ , $C=2^8$ )	0,17	0,38	5,13	11,28	5,14	11,28
Máquinas de vetor de suporte ( $\epsilon=0.5$ , $\gamma=2^{-1.5}$ , $C=2^{8.5}$ )	0,15	0,32	5,22	11,47	5,22	11,47
Máquinas de vetor de suporte ( $\gamma=2^{-1.5}$ , $C=2^{8.5}$ )	0,24	0,54	5,25	11,53	5,25	11,54
Máquinas de vetor de suporte ( $\gamma=2^{-1.5}$ , $C=2^8$ )	0,24	0,54	5,25	11,53	5,25	11,54
Máquinas de vetor de suporte ( $\epsilon=0.5$ , $\gamma=2^{-1.5}$ , $C=2^8$ )	0,27	0,59	5,50	12,08	5,51	12,09
Máquinas de vetor de suporte ( $\gamma=2^{-1.5}$ , $C=2^8$ )	0,34	0,75	5,54	12,16	5,55	12,18
Máquinas de vetor de suporte ( $\gamma=2^{-1.5}$ , $C=2^8$ )	0,34	0,75	5,54	12,16	5,55	12,18
Máquinas de vetor de suporte ( $\epsilon=0.5$ , $\gamma=2^{-2.0}$ , $C=2^{8.5}$ )	0,17	0,36	5,58	12,26	5,58	12,27
Máquinas de vetor de suporte ( $\gamma=2^{-2.0}$ , $C=2^{8.5}$ )	0,17	0,37	5,60	12,30	5,60	12,30
Máquinas de vetor de suporte ( $\epsilon=0.5$ , $\gamma=2^{-2.0}$ , $C=2^8$ )	0,17	0,37	5,68	12,47	5,68	12,48
Máquinas de vetor de suporte ( $\gamma=2^{-2.0}$ , $C=2^8$ )	0,16	0,35	5,72	12,56	5,72	12,56
Máquinas de vetor de suporte ( $\epsilon=0.5$ , $\gamma=2^{-2.5}$ , $C=2^{8.5}$ )	0,14	0,31	5,81	12,77	5,82	12,77
Máquinas de vetor de suporte ( $\gamma=2^{-2.5}$ , $C=2^{8.5}$ )	0,16	0,35	5,85	12,84	5,85	12,85
Máquinas de vetor de suporte ( $\epsilon=0.5$ , $\gamma=2^{-2.5}$ , $C=2^8$ )	0,16	0,36	5,92	13,01	5,93	13,02
Máquinas de vetor de suporte ( $\gamma=2^{-2.5}$ , $C=2^8$ )	0,18	0,39	5,96	13,10	5,97	13,11
Máquinas de vetor de suporte ( $\gamma=2^{-0.5}$ , $C=2^{7.5}$ )	0,33	0,72	6,52	14,33	6,53	14,35

FONTE: O autor (2020).

Os valores de biomassa observado e preditos podem ser visualizados na Figura 28, para os 4 melhores estimadores usando máquinas de vetores de suporte.

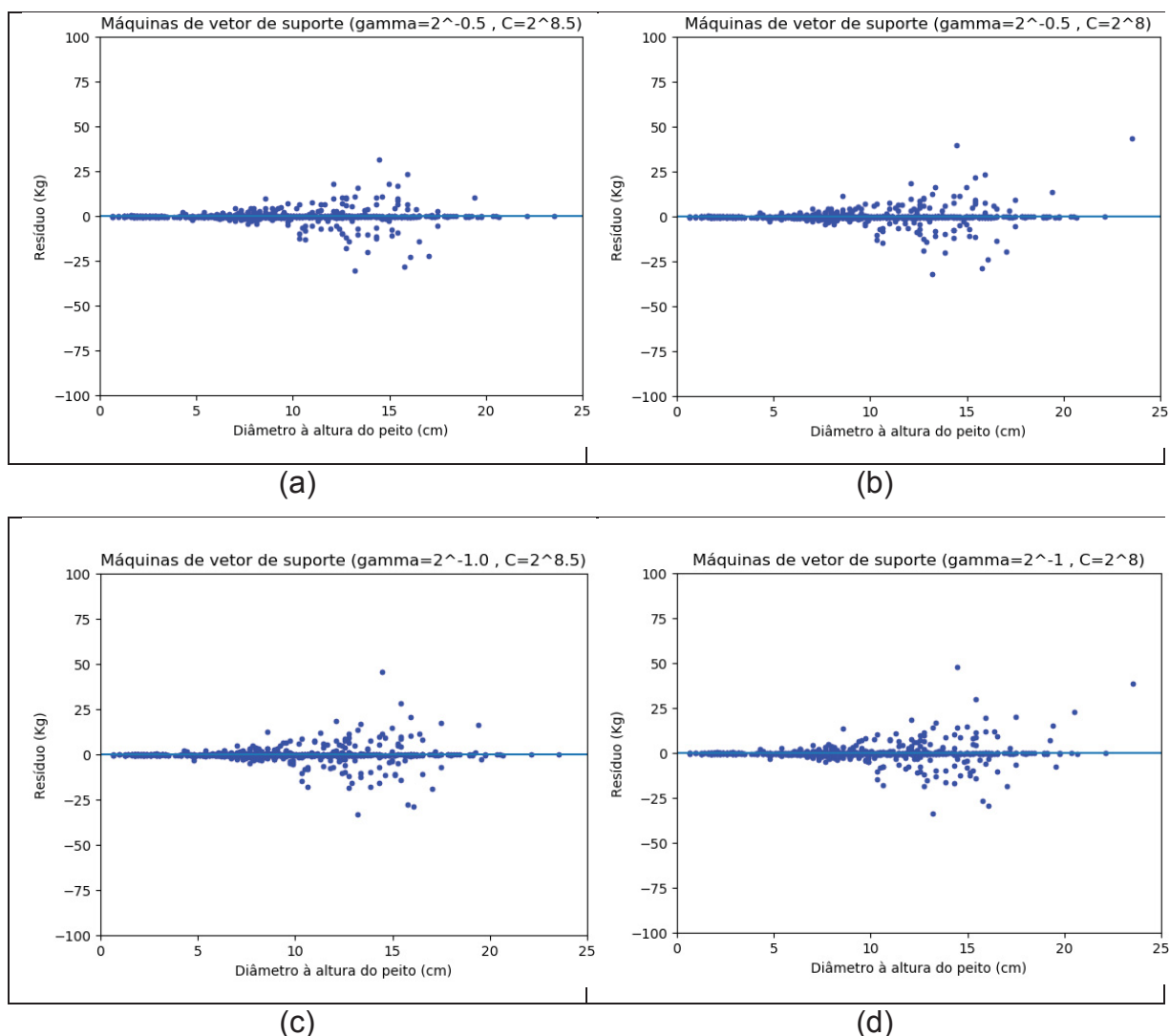
FIGURA 28—VALORES OBSERVADOS DA VARIÁVEL BIOMASSA DE ÁRVORES DE ACÁCIA NEGRA E ESTIMADOS POR MÁQUINAS DE VETORES DE SUPORTE.



FONTE: O autor (2020).

Na Figura 29 observa-se os resíduos (em Kg) para os 4 estimadores com melhores resultados usando máquinas de vetores de suporte. Foi observado que os 4 melhores resultados apresentaram *overfitting*, o que acarretaria em uma predição ruim conforme apresentado na Tabela 15. Para solucionar esse problema ao invés de utilizar somente os 4 estimadores selecionados, foram utilizados todos os estimadores obtidos no refinamento para observar as informações da base de dados de validação.

FIGURA 29– GRÁFICO DE RESÍDUOS (em Kg) DA VARIÁVEL BIOMASSA DE ÁRVORES DE ACÁCIA NEGRA MODELADA POR MÁQUINAS DE VETORES DE SUPORTE



FONTE: O autor (2020).

#### 4.3.4.1 Predição da base de dados para validação 98/2012

Com o problema de *overfitting* apresentado nos melhores ajustes, optou-se por uma abordagem diferenciada das apresentadas nos demais algoritmos de aprendizagem de máquina, sendo assim, optou-se por estimar o viés, precisão e acurácia de todos os estimadores de máquinas de vetores de suporte apresentados no refinamento. Dessa forma foi realizado o processo inverso, pegou-se os quatro melhores estimadores para a base de dados 98/2012, e observou-se os resultados para a base de dados 670/2014. Os resultados dessas estatísticas estão apresentados na Tabela 15.

TABELA 15 – RESULTADOS DE VIÉS, PRECISÃO E ACURÁCIA PARA OS ESTIMADORES DA VARIÁVEL BIOMASSA DE ÁRVORES DE ACÁCIA NEGRA UTILIZANDO MODELOS DE MÁQUINAS DE VETORES DE SUPORTES, PARA A BASE DE DADOS DE **VALIDAÇÃO 98/2012**.

Estimador	$\bar{e}$	$\bar{e}_{\%}$	$S_e$	$S_{e\%}$	$m_x$	$m_{x\%}$
Máquinas de vetores de suporte (epsilon=0.5, gamma=2 <sup>-1</sup> .5 , C=2 <sup>^</sup> 8.5)	12,2	16,3		18,3		24,5
	3	4	13,72	2	18,38	5
Máquinas de vetores de suporte (epsilon=0.5, gamma=2 <sup>-1</sup> , C=2 <sup>^</sup> 8.5)	12,6	16,8		19,2		25,5
	0	3	14,42	6	19,15	8
Máquinas de vetores de suporte (epsilon=0.5, gamma=2 <sup>-0</sup> .5 , C=2 <sup>^</sup> 7.5)	12,8	17,2		19,7		26,2
	9	2	14,82	9	19,64	3
Máquinas de vetores de suporte (epsilon=0.5, gamma=2 <sup>-1</sup> .5 , C=2 <sup>^</sup> 7.5)	13,3	17,8		21,6		28,0
	8	7	16,21	5	21,01	7
Máquinas de vetores de suporte (epsilon=0.5, gamma=2 <sup>-1</sup> , C=2 <sup>^</sup> 7.5)	14,4	19,2		25,0		31,5
	2	6	18,71	0	23,62	6
Máquinas de vetores de suporte (epsilon=0.5, gamma=2 <sup>-0</sup> .5 , C=2 <sup>^</sup> 8)	15,8	21,1		29,0		35,9
	6	8	21,73	2	26,90	3
Máquinas de vetores de suporte (epsilon=0.5, gamma=2 <sup>-1</sup> .5 , C=2 <sup>^</sup> 8)	18,2	24,4		35,6		43,2
	7	1	26,69	6	32,35	1
Máquinas de vetores de suporte (epsilon=0.5, gamma=2 <sup>-2</sup> .5 , C=2 <sup>^</sup> 8)	17,5	23,5		40,8		47,1
	9	0	30,61	9	35,30	6
Máquinas de vetores de suporte (epsilon=0.5, gamma=2 <sup>-2</sup> .5 , C=2 <sup>^</sup> 8.5)	17,1	22,9		41,3		47,3
	6	2	30,98	8	35,41	1
Máquinas de vetores de suporte (epsilon=0.5, gamma=2 <sup>-2</sup> .5 , C=2 <sup>^</sup> 7.5)	18,1	24,1		40,6		47,3
	1	9	30,46	9	35,44	4
Máquinas de vetores de suporte (gamma=2 <sup>-2</sup> .5 , C=2 <sup>^</sup> 8)	17,6	23,5		41,2		47,5
	5	8	30,87	3	35,56	0
Máquinas de vetores de suporte (gamma=2 <sup>-2</sup> .5 , C=2 <sup>^</sup> 7.5)	18,1	24,1		40,9		47,5
	0	7	30,65	5	35,60	5
Máquinas de vetores de suporte (gamma=2 <sup>-2</sup> .5 , C=2 <sup>^</sup> 8.5)	17,2	23,0		41,7		47,6
	3	1	31,27	7	35,70	9
Máquinas de vetores de suporte (epsilon=0.5, gamma=2 <sup>-0</sup> .5 , C=2 <sup>^</sup> 8.5)	20,7	27,6		42,3		50,5
	2	8	31,69	4	37,86	8
Máquinas de vetores de suporte (epsilon=0.5, gamma=2 <sup>-1</sup> , C=2 <sup>^</sup> 8)	21,0	28,1		42,8		51,2
	5	2	32,09	7	38,38	7
Máquinas de vetores de suporte (epsilon=0.5, gamma=2 <sup>-1</sup> .0 , C=2 <sup>^</sup> 8.5)	18,5	24,7		47,3		53,4
	2	4	35,43	3	39,98	0
Máquinas de vetores de suporte (gamma=2 <sup>-1</sup> .0 , C=2 <sup>^</sup> 8.5)	18,7	25,0		47,3		53,5
	2	1	35,42	1	40,06	1
Máquinas de vetores de suporte (epsilon=0.5, gamma=2 <sup>-1</sup> .0 , C=2 <sup>^</sup> 8)	19,0	25,3		47,2		53,5
	0	8	35,33	0	40,12	9
Máquinas de vetores de suporte (gamma=2 <sup>-1</sup> .0 , C=2 <sup>^</sup> 8)	19,0	25,4		47,2		53,7
	7	8	35,39	8	40,20	1
Máquinas de vetores de suporte (epsilon=0.5, gamma=2 <sup>-2</sup> .0 , C=2 <sup>^</sup> 7.5)	19,6	26,2		47,5		54,3
	6	6	35,58	3	40,65	1

FONTE: O autor (2020).

Para facilitar a visualização dos resultados da modelagem da biomassa por meio das técnicas de aprendizagem de máquina, propôs-se apresentar os resultados de viés, precisão e acurácia dos dados originais e os resultados escolhidos para os dados de validação, com o intuito de observar se houve discrepância entre eles. Esses resultados estão apresentados na Tabela 16.

TABELA 16 – COMPARAÇÃO DE VIÉS, PRECISÃO E ACURÁCIA PARA OS MELHORES ESTIMADORES DE MÁQUINA DE VETORES DE SUORTE PARA A BASE DE DADOS DE ESTIMAÇÃO E DE VALIDAÇÃO, EM RELAÇÃO AOS RESULTADOS OBTIDOS NA BASE DE DADOS UTILIZADA NA CRIAÇÃO DOS ESTIMADORES.

Estimador	$\bar{e}$	$\bar{e}_{\%}$	$S_e$	$S_{e\%}$	$m_x$	$m_{x\%}$
Máquinas de vetor de suporte (gamma=2 <sup>-0</sup> .5 , C=2 <sup>^</sup> 8.5)	0,0	0,0			4,0	
	1	2	4,01	8,81	1	8,81
Máquinas de vetor de suporte (gamma=2 <sup>-0</sup> .5 , C=2 <sup>^</sup> 8)	0,1	0,3		10,0	4,5	10,0
	4	0	4,57	5	8	5
Máquinas de vetor de suporte (gamma=2 <sup>-1</sup> .0 , C=2 <sup>^</sup> 8.5)	0,0	0,1		10,2	4,6	10,2
	9	9	4,67	6	7	6
Máquinas de vetor de suporte (gamma=2 <sup>-1</sup> , C=2 <sup>^</sup> 8)	0,1	0,3		11,2	5,1	11,2
	7	8	5,13	8	4	8
Máquinas de vetor de suporte (epsilon=0.5, gamma=2 <sup>-1</sup> .5 , C=2 <sup>^</sup> 8.5)	0,1	0,3		11,4	5,2	11,4
	5	2	5,22	7	2	7

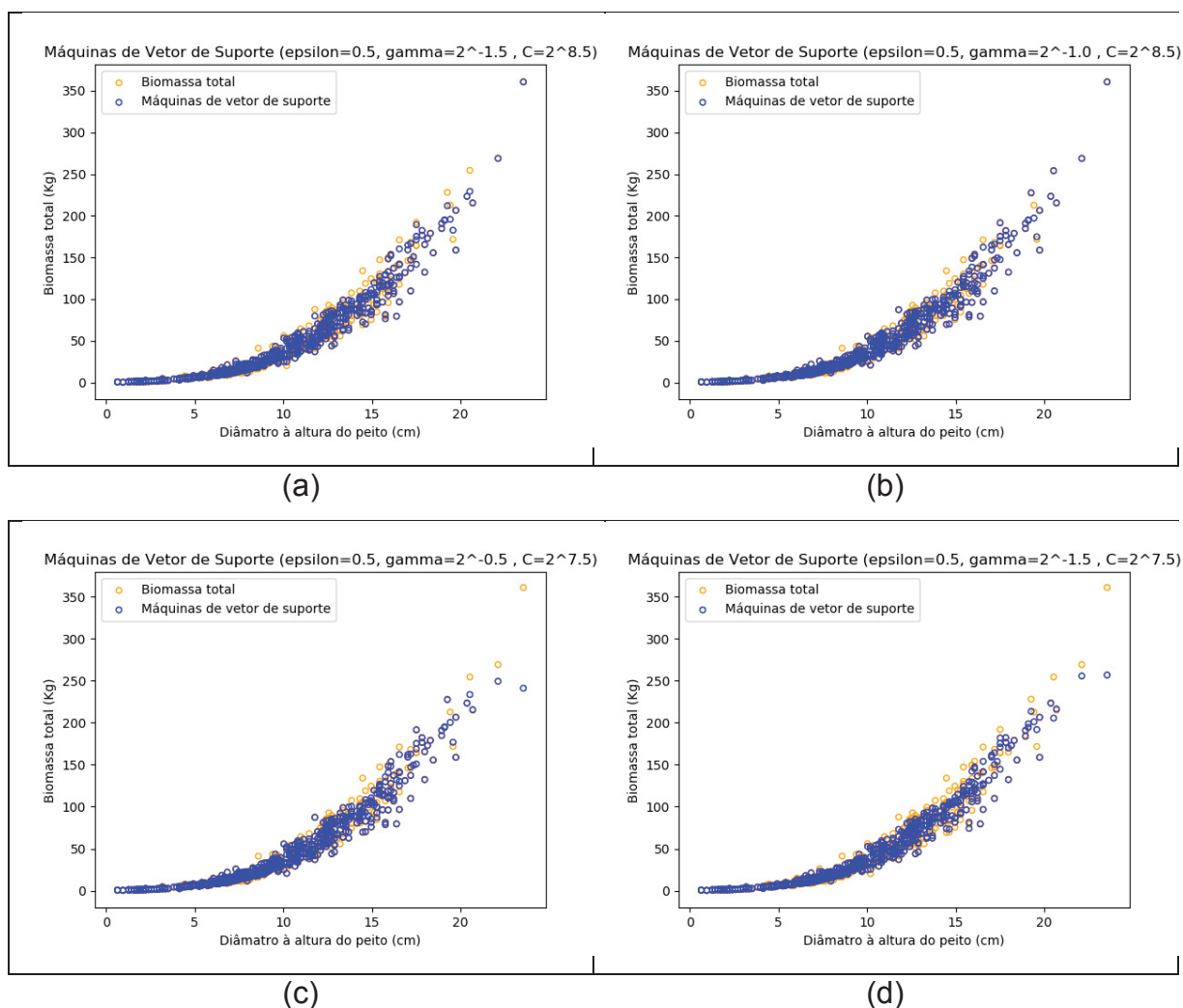
Máquinas de vetor de suporte (epsilon=0.5, gamma=2 <sup>-1</sup> , C=2 <sup>8.5</sup> )	0,3	0,8		18,4	8,4	18,4
	7	1	8,41	7	2	9
Máquinas de vetor de suporte (epsilon=0.5, gamma=2 <sup>-0,5</sup> , C=2 <sup>7.5</sup> )	0,2	0,6		19,0	8,6	19,0
	9	4	8,66	2	6	3
Máquinas de vetor de suporte (epsilon=0.5, gamma=2 <sup>-1,5</sup> , C=2 <sup>7.5</sup> )	0,6	1,4		20,9	9,5	20,9
	5	2	9,52	1	4	5

FONTE: O autor (2020).

LEGENDA: Primeiros quatro estimadores: melhores resultados para a base de dados 670/2014. Outros quatro estimadores: resultados observados na base de dados 98/2012, para a base de dados de 670/2014.

Todos os gráficos foram refeitos para a nova escolha de modelos e estão apresentados na Figura 30.

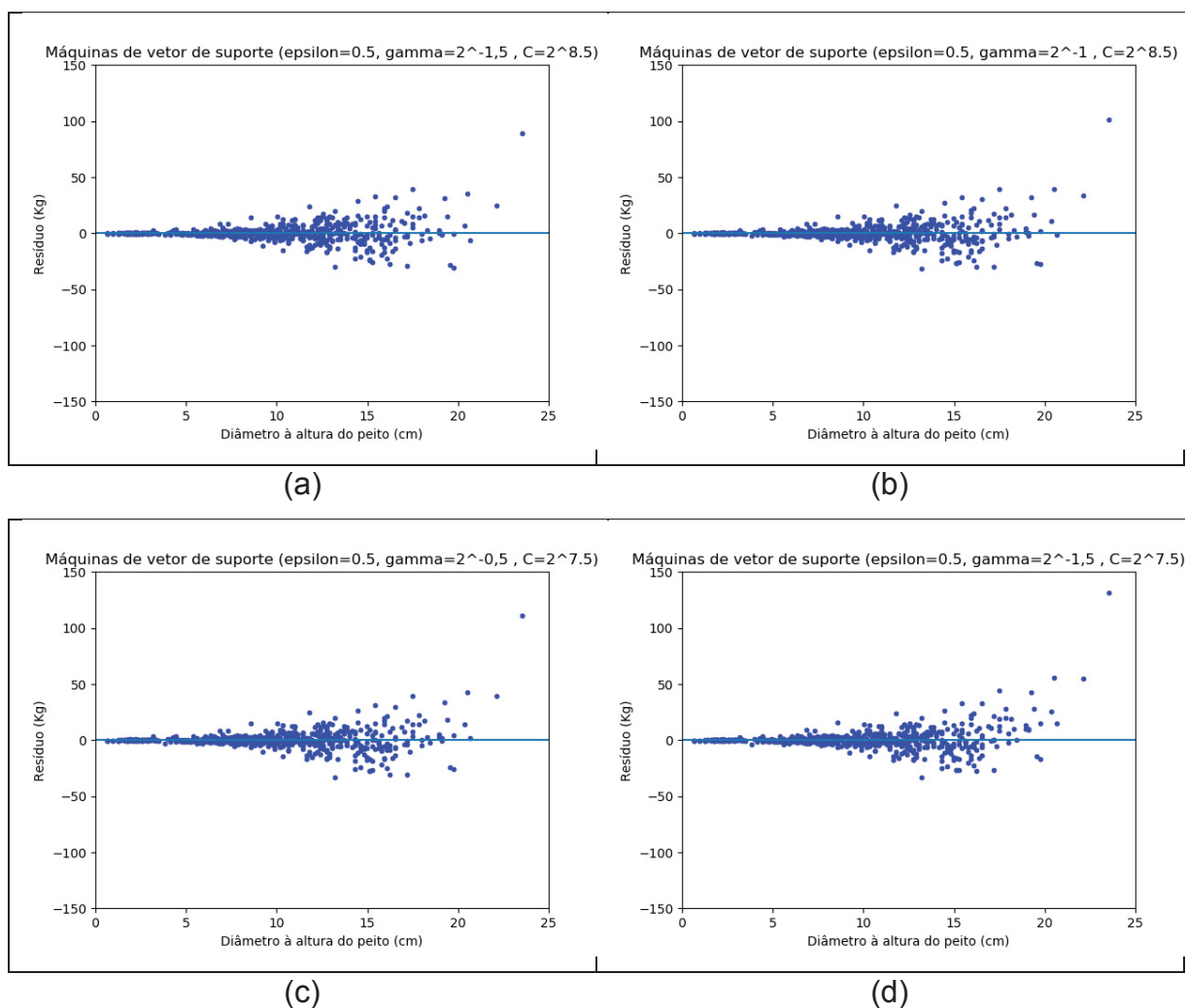
FIGURA 30– GRÁFICO DE DISPERSÃO DOS ESTIMADORES POR MÁQUINA DE VETORES DE SUORTE EM COMPARAÇÃO COM A BIOMASSA TOTAL.



FONTE: O autor (2020).

O gráfico de resíduos (Kg) dos dados originais também foram refeitos para os quatro estimadores com melhores resultados nos dados de validação, Figura 31.

FIGURA 31– GRÁFICO DE RESÍDUOS DOS ESTIMADORES POR MÁQUINA DE VETORES DE SUORTE EM COMPARAÇÃO COM A BIOMASSA TOTAL.



FONTE: O autor (2020).

#### 4.4 COMPARAÇÃO DAS TÉCNICAS DE APRENDIZAGEM DE MÁQUINA APLICADAS NA MODELAGEM DE BIOMASSA



#### 4.4.1 Avaliação do viés, precisão e acurácia

Para melhor observar e comparar as técnicas de aprendizagem de máquina foi realizado um resumo dos valores de viés, precisão e acurácia usando a base de dados utilizada no treinamento, 670/2014, (Tabela 17) e a base de dados utilizado na validação (Tabela 18).

**TABELA 17 – COMPARAÇÃO DE VIÉS, PRECISÃO E ACURÁCIA PARA OS MELHORES ESTIMADORES DE BIOMASSA DE ÁRVORES DE ACÁCIA-NEGRA MODELADA POR MEIO DE TÉCNICAS DE APRENDIZAGEM DE MÁQUINA, PARA A BASE DE DADOS DE TREINAMENTO 670/2014.**

Estimador	$\bar{e}$	$\bar{e}_{\%}$	$S_e$	$S_{e\%}$	$m_x$	$m_{x\%}$
k-Vizinhos mais próximos (n_neighbors=5, weights = "distance")	-3,71E-17	-8,14E-17	0,73	1,60	0,73	1,60
k-Vizinhos mais próximos (n_neighbors=6, weights = "distance")	-3,71E-17	-8,14E-17	0,73	1,60	0,73	1,60
k-Vizinhos mais próximos (n_neighbors=7, weights = "distance")	-3,71E-17	-8,14E-17	0,73	1,60	0,73	1,60
k-Vizinhos mais próximos (n_neighbors=8, weights = "distance")	-3,71E-17	-8,14E-17	0,73	1,60	0,73	1,60
Floresta aleatória (n_estimators=60, oob_score="True")	0,03	0,06	3,62	7,96	3,62	7,96
Floresta aleatória (n_estimators=100, oob_score="True")	-0,02	-0,04	3,66	8,05	3,66	8,05
Floresta aleatória (n_estimators=450, oob_score="True")	0,01	0,02	3,69	8,11	3,69	8,11
Floresta aleatória (n_estimators=300, oob_score="True")	-0,05	-0,11	3,71	8,14	3,71	8,14
Máquinas de vetor de suporte (epsilon=0.5, gamma=2 <sup>-1.5</sup> , C=2 <sup>8.5</sup> )	0,15	0,32	5,22	11,47	5,22	11,47
Redes Neurais Artificiais (momentum=0.6, learning_rate_init=0.03)	-0,18	-0,39	7,01	15,40	7,01	15,40
Redes Neurais Artificiais (momentum=0.6, learning_rate_init=0.30)	-0,07	-0,14	7,21	15,83	7,21	15,84
Redes Neurais Artificiais (momentum=0.4, learning_rate_init=0.09)	-0,05	-0,11	7,44	16,35	7,44	16,35
Redes Neurais Artificiais (momentum=0.3, learning_rate_init=0.25)	0,18	0,40	7,49	16,46	7,49	16,46
Máquinas de vetor de suporte (epsilon=0.5, gamma=2 <sup>-1</sup> , C=2 <sup>8.5</sup> )	0,37	0,81	8,41	18,47	8,42	18,49
Máquinas de vetor de suporte (epsilon=0.5, gamma=2 <sup>-0.5</sup> , C=2 <sup>7.5</sup> )	0,29	0,64	8,66	19,02	8,66	19,03
Máquinas de vetor de suporte (epsilon=0.5, gamma=2 <sup>-1.5</sup> , C=2 <sup>7.5</sup> )	0,65	1,42	9,52	20,91	9,54	20,95

FONTE: O autor (2020).

**TABELA 18 – COMPARAÇÃO DE VIÉS, PRECISÃO E ACURÁCIA PARA OS MELHORES ESTIMADORES DE BIOMASSA DE ÁRVORES DE ACÁCIA-NEGRA MODELADA POR MEIO DE TÉCNICAS DE APRENDIZAGEM DE MÁQUINA, PARA A BASE DE DADOS DE VALIDAÇÃO 98/2012.**

Estimador	$\bar{e}$	$\bar{e}_{\%}$	$S_e$	$S_{e\%}$	$m_x$	$m_{x\%}$
Floresta aleatória (n_estimators=100, oob_score="True")	9,96	13,31	10,89	14,55	14,76	19,72
Floresta aleatória (n_estimators=300, oob_score="True")	9,97	13,32	10,97	14,65	14,83	19,80
Floresta aleatória (n_estimators=60, oob_score="True")	10,36	13,85	11,12	14,86	15,20	20,31
Floresta aleatória (n_estimators=450, oob_score="True")	10,23	13,67	11,32	15,12	15,26	20,38
Redes Neurais Artificiais (momentum=0.6, learning_rate_init=0.03)	10,31	13,77	12,76	17,04	16,4	21,91
Redes Neurais Artificiais (momentum=0.4, learning_rate_init=0.09)	11,7	15,63	12,92	17,25	17,43	23,28
Máquinas de vetores de suporte (epsilon=0.5, gamma=2 <sup>-1.5</sup> , C=2 <sup>8.5</sup> )	12,23	16,34	13,72	18,32	18,38	24,55
Redes Neurais Artificiais (momentum=0.6, learning_rate_init=0.30)	11,64	15,56	14,47	19,33	18,58	24,81
Máquinas de vetores de suporte (epsilon=0.5, gamma=2 <sup>-1</sup> , C=2 <sup>8.5</sup> )	12,6	16,83	14,42	19,26	19,15	25,58

Máquinas de vetores de suporte (epsilon=0.5, gamma=2 <sup>-0,5</sup> , C=2 <sup>7,5</sup> )	12,89	17,22	14,82	19,79	19,64	26,23
Máquinas de vetores de suporte (epsilon=0.5, gamma=2 <sup>-1,5</sup> , C=2 <sup>7,5</sup> )	13,38	17,87	16,21	21,65	21,01	28,07
k-Vizinhos mais próximos (n_neighbors=7, weights = "distance")	13,95	18,64	18,63	24,89	23,28	31,09
k-Vizinhos mais próximos (n_neighbors=6, weights = "distance")	14,05	18,76	18,87	25,21	23,53	31,43
k-Vizinhos mais próximos (n_neighbors=8, weights = "distance")	14,34	19,16	19,15	25,58	23,93	31,96
k-Vizinhos mais próximos (n_neighbors=5, weights = "distance")	14,18	18,94	19,68	26,28	24,25	32,40
Redes Neurais Artificiais (momentum=0.3, learning_rate_init=0.25)	14,69	19,63	20,09	26,84	24,89	33,25

FONTE: O autor (2020).

Os valores do viés relativo estão no intervalo de  $-8,14E^{-17}$  a 1,42%, a precisão teve sua variação entre 1,60% a 20,91% e a acurácia entre 1,60% e 20,95%. Ao comparar-se os resultados com a literatura (discutidos na sequência), observa-se que os resultados foram muitas vezes melhores que os resultados por modelos de regressão lineares, não-lineares, inclusive aos de aprendizagem de máquina.

Já publicados, Rezende et al. (2006) afirmam que o viés em modelagem de biomassa para o cerrado variou entre 26,34% a 33,76%. Silva (2014) descreve que o viés de biomassa de *Eucalyptus grandis* foi de 0,61 Mg. Nicoletti et al. (2015) apresentaram um viés negativo de 15% para biomassa do fuste de remanescentes de floresta ombrófila mista, ou seja, houve uma subestimativa de tal magnitude. Schneider et al. (2005) descreveram um viés de 4,4% para determinar o estoque de biomassa em povoamentos de acácia-negra. Kuyah et al. (2014) apresentaram resultados para florestas com predominância de *Acacia mearnsii*, *Eucalyptus spp.* e *Markhamia lutea* de modo que o viés encontrado para seus modelos de biomassa variaram de -34,9% a 29,7%. Esses valores demonstram que os resultados encontrados nesse trabalho são condizentes com a literatura e em certos casos superiores aos obtidos pelos autores anteriormente mencionados.

Schilowski et al. (2015) observaram que a precisão para biomassa de Álamo variou de 10,40% a 14,14%, utilizando-se aprendizagem de máquina. Valerio et al. (2007) descreveram que seus modelos para biomassa de uva-do-Japão tiveram precisão variando de 9,77% a 35,02%. Sanquetta et al. (2016b) apresentaram valores de precisão com 12,81% a 15,84% para estimar biomassa arbórea de acácia-negra.

Behling (2009) descreveu uma acurácia de 26,9% e 14,5% para biomassa de folha e fuste de indivíduos de Teca na região de Tangará na serra do Mato Grosso - Brasil. Palma (2014) descreveu que a acurácia relativa para biomassa de *Acacia mangium* Wild. foi de 0,0502. Poudel et al. (2015) descrevem que a estimativa de biomassa acima do solo de *Pseudotsuga menziesii* teve acurácia 2,3% a 57,7% para

modelos de biomassa lineares e não-lineares. Ao analisar a literatura pode-se concluir que os estimadores obtidos nesse trabalho estão dentro do observado por outros autores, sendo muitas vezes melhores do que os apresentados pelas demais literaturas.

Montaño (2016) utilizou em sua pesquisa parte do banco de dados 670/2014. O autor estimou a biomassa dos indivíduos de acácia-negra e pelos métodos de regressão (*Kopecky & Gehrhard, Spurr e Schumacher & Hall*) bem como técnicas de aprendizagem de máquina (redes neurais artificiais, máquina de vetores de suporte e florestas aleatórias) onde obteve acurácia relativa (%): 22,72, 19,46, 19,66, 19,80, 17,68 e 22,10, respectivamente. Ao comparar os resultados do autor com os apresentados na Tabela 17, tem-se 13 estimadores de 16 dos apresentados no presente trabalho com resultados superiores.

Contudo, ao analisar-se os valores de validação pode-se constatar que seus resultados são inferiores a outras técnicas de aprendizagem de máquina quando o quesito é estimar outros bancos de dados, visto que o viés aumenta para 17,22%, a precisão piora para 19,79% e acurácia alcança 26,23%. Em seguida os resultados de *k*-vizinhos mais próximos com a métrica *Chebyshev* e com 2 vizinhos tiveram valores de viés entre 0,08 a 0,21%, com precisão e acurácia variando de 11,95% a 12,66%.

Em contrapartida, o melhor resultado para os dados de validação foi a floresta aleatória com 60 árvores e *min\_sample\_leaf* com valor de 2. A mesma apresentou viés de 12,42%, precisão de 15,34% e acurácia foi de 19,74%. Os resultados para avaliação dos estimadores foram: viés de -0,14%, precisão e acurácia de 13,63%.

Ao avaliar o resultado de validação notou-se que os melhores resultados foram as floresta aleatória, variando o viés de 12,42% a 13,13%, precisão de 14,78% a 15,65% e acurácia de 19,74% a 20,92%. Os piores resultados foram dos quatro modelos de máquinas de vetores de suporte e a rede neural artificial com momentum de 0,3 e *learning\_rate\_init* de 0,25, com variação de viés de 17,32% a 20,43%, precisão de 24,33% a 28,64% e acurácia com valores entre 29,87% a 35,17%.

Ao analisar-se todos os resultados apresentados, observou-se que muitas vezes, resultados excelentes para estimar dados podem apresentar pouca flexibilidade e criar resultados ruins para novos dados. A decisão do melhor resultado foi baseada em observar os estimadores, de modo que o melhor deveria ser bom tanto na base de dados 670/2014 quanto nos dados da validação 98/2012. A floresta aleatória com 100 árvores, *min\_sample\_init* de 1 e *oob\_score true* revelou melhor

resultado de árvores de decisão para a base de dados 670/2014 e o terceiro melhor resultado para os dados de validação.

#### 4.4.2 Hiperparâmetros dos modelos de aprendizagem de máquina

Os hiperparâmetros, para cada técnica de aprendizagem de máquina, avaliados como os de melhores resultados para a modelagem da biomassa de acácia-negra estão sumarizados na Tabela 19. Foram também destacados alguns desses para outras espécies por meio da literatura. Essas informações são úteis, pelo menos inicialmente, dos hiperparâmetros que podem ser otimizados em pesquisas futuras.

TABELA 19 – HIPERPARÂMETROS DOS ALGORITMOS DE APRENDIZAGEM DE MÁQUINA AJUSTADOS PARA A MODELAGEM DA VARIÁVEL BIOMASSA PARA A ESPÉCIE ACÁCIA-NEGRA E TAMBÉM PARA OUTRAS.

Estimador	Hiperparâmetro	Autor
Floresta aleatória	100 árvores, oob_score="True"	Martins (2020)
Máquinas de vetor de suporte	epsilon=0.5, gamma=2 <sup>-1.5</sup> , C=2 <sup>8.5</sup>	Martins (2020)
k-Vizinhos mais próximos	n_neighbors=7, weights = "distance"	Martins (2020)
Redes neurais artificiais	momentum=0.6, learning_rate_init=0.03, hidden_layer_sizes=(15, 15)	Martins (2020)
Redes neurais artificiais	momentum=0.2, learning_rate_init=0.5, hidden_layer_sizes=(18)	Montaño (2016)
Floresta aleatória	150 árvores	Montaño (2016)
Máquinas de vetor de suporte	gamma=0.1	Montaño (2016)
Floresta aleatória	100 árvores	Schikoski et al. (2018)
k-Vizinhos mais próximos	n_neighbors=20, metric="euclidean"	Schikoski et al. (2018)
Redes neurais artificiais	momentum=0.4	Schikoski et al. (2018)
Máquinas de vetor de suporte	epsilon=0.5, gamma=2 <sup>-4.8</sup> , C=2 <sup>4.5</sup>	Costa Filho (2019)
Floresta aleatória	390 árvores, <i>min_sample_leaf</i> com valor de 2	Costa Filho (2019)
k-Vizinhos mais próximos	n_neighbors=16, metric="manhattan"	Costa Filho (2019)
Redes neurais artificiais	hidden_layer_sizes=(13), activation="maxout"	Costa Filho (2019)
Redes neurais artificiais	hidden_layer_sizes=(9)	Schuh (2019)
Máquinas de vetor de suporte	epsilon=0.1, gamma=0,01694915254, kernel="poly", degree=3	Schuh (2019)
Floresta aleatória	1.000 árvores <i>min_sample_leaf</i> com valor de 29	Schuh (2019)

FONTE: O autor (2020).

A quantidade de combinação de parâmetros é ampla para as técnicas de aprendizagem de máquina, gerando modelos distintos para a modelagem de uma única variável, como nesta pesquisa (biomassa da espécie acácia-negra). Conforme Tabela 19, diferentes configurações podem ser obtidas e não há um consenso, ainda,

sobre qual é a melhor. Ao contrário dos modelos de regressão, que após muitos anos de pesquisa, já se tem uma ideia dos melhores modelos.

Sendo assim, as florestas aleatórias foram os melhores resultados para aprendizagem de máquinas, especificamente a floresta com 100 árvores e classificação *out-of-bag*, e serão comparadas com a técnica de testemunha.

#### 4.5 TÉCNICA TESTEMUNHA

Os pacotes utilizados para desenvolver a programação para estimar a biomassa por meio das equações de regressão foram: *DataFrames*, *XLSX*, *Statistics* e *PyPlot*. No Algoritmo 16, a base de dados 670/2014 foi particionada em função do estrato de idade, definido na Tabela 1 (pág 54), para aplicação das regressões de biomassa para esses estratos.

ALGORITMO16 – REPRESENTAÇÃO DA PROGRAMAÇÃO DE EQUAÇÕES NÃO-LINEARES PARA ESTIMAR A BIOMASSA EM LINGUAGEM DE PROGRAMAÇÃO *JULIA*.

```
julia> # Definição dos parâmetros
julia> MJ = dados[dados.i.<= 1,:]
julia> J = dados[1.<dados.i.<=2,:]
julia> MI = dados[2.<dados.i.<=4,:]
julia> MA = dados[4.<dados.i.<=7,:]
julia> M = dados[9.<dados.i.<11,:]
julia> # Equações
julia> function Equacao(dados)
    Ymj=(0.205422.*((MJ.d.^2).*MJ.h).^0.459328).+(0.17493*((MJ.d.^2).*MJ.h).
    ^0.506569)
    Yj=(0.033033.*((J.d.^2).*J.h).^0.827421).+(0.043660*((J.d.^2).*J.h).^0.8
    67317)
    Ymi=(0.003976.*((MI.d.^2).*MI.h).^1.056877).+(0.051761*((MI.d.^2).*MI.h)
    .^0.863215)
    Yma=(0.008456.*((MA.d.^2).*MA.h).^0.931885).+(0.087331*(MA.d.^2.146802).
    *(MA.h.^0.353867))
    Ym=(0.000059.*((M.d.^2).*M.h).^1.513338).+(0.146367*(M.d.^2.013462).*(M.
    h.^0.350896))
    yhat = vcat(Ymj, Yj, Ymi, Yma, Ym)
    end
julia> # Obter estimativa
```

```
julia> Equacao(dados) # nesse caso utiliza-se a expressão dados pois a criação
da função gera um arquivo completo para todos os valores da base de dados,
sendo assim não se utiliza a expressão de uma das classes.
```

FONTE: O autor (2020).

Os resultados do viés, precisão e acurácia para os modelos de regressão de biomassa estão descritos na Tabela 20.

TABELA 20 – RESULTADOS DE VIÉS, PRECISÃO E ACURÁCIA PARA OS ESTIMADORES DA VARIÁVEL BIOMASSA DE ÁRVORES DE ACÁCIA-NEGRA UTILIZANDO MODELOS DE REGRESSÃO, USANDO A BASE DE DADOS DE TREINAMENTO 670/2014.

Estimador	$\bar{e}$	$\bar{e}_{\%}$	$S_e$	$S_{e\%}$	$m_x$	$m_{x\%}$
Equações WNSUR <sup>6</sup>	1,8311	4,02	3,2636	7,1690	3,7422	8,22

FONTE: O autor (2020).

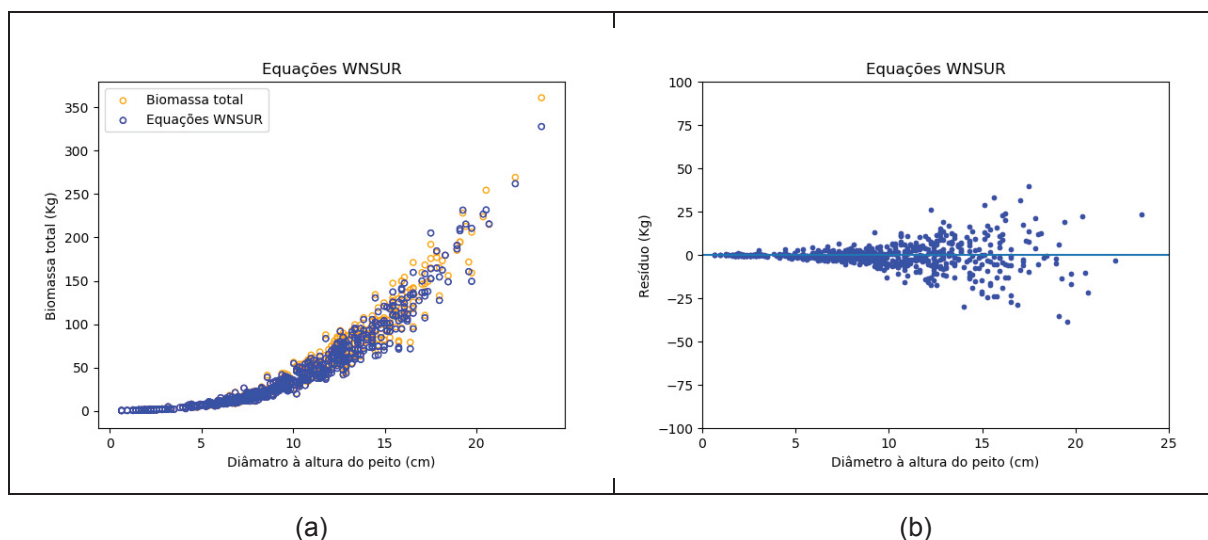
Os valores observados, preditos e também os resíduos (em kg) podem ser visualizados na Figura 32. Assim como observado nas técnicas de aprendizagem de máquina, o erro de estimativa (em unidades absolutas) aumenta nas árvores maiores e os resíduos são heteroscedásticos.

Ressalta-se que os resíduos da maneira que foram calculados são heteroscedásticos, ou seja, conforme aumenta o tamanho da árvore aumenta também a variância nos resíduos. Para resolver esse problema na análise de regressão, Behling et al. (2019) ajustaram essas equações de maneira ponderada. Assim, para obtenção de resíduos homoscedásticos, os resíduos obtidos na Figura 20 devem ser multiplicados pelos resultados das funções de pesos apresentadas pelos autores. Essa figura não foi construída, pois acredita-se que essa não seria importante para comparação com as técnicas de aprendizagem de máquina, dada a não necessidade de correção da heteroscedasticidade dos resíduos.

---

<sup>6</sup>Weighted-Nonlinear Seemingly Unrelated Regression

FIGURA 32– VALORES OBSERVADOS E ESTIMADOS DA VARIÁVEL BIOMASSA DE ÁRVORES DE ACÁCIA-NEGRA E RESÍDUOS (em kg) OBTIDOS POR MEIO DE MODELOS DE REGRESSÃO.



FONTE: O autor (2020).

A dispersão dos resíduos de maneira heterogênea em relação às variáveis aleatórias utilizadas nas funções de biomassa, ou em relação à própria variável biomassa, indica que a variância não é constante. A heteroscedasticidade dos resíduos é comum na modelagem da biomassa arbórea, podendo também ser observada nos trabalhos de Sanquetta et al. (2015) e Zhao et al. (2015), entre outros. Normalmente, resíduos homoscedásticos na biomassa são observados quando a amostra utilizada é pequena, o que dificulta observar essa relação.

Na modelagem por análise de regressão, a avaliação da heterogeneidade dos resíduos é importante, porque tem consequência nas propriedades do estimador de regressão por mínimos quadrados. Dada a condição de heterogeneidade, os estimadores de mínimos quadrados ainda não são enviesados (a média é apropriadamente estimada), mas são inefficientes (as estimativas das variâncias são enviesadas) invalidando, por consequência, os testes de significância F e t na análise de regressão, conforme descrito por Greene (2008).

Para contornar o problema da heteroscedasticidade, a variável biomassa pode ser transformada ou a regressão pode ser ajustada com a utilização de pesos, assim como realizado pelos autores que ajustaram as equações de biomassa. Na análise de regressão, além do pesquisador se preocupar com qual modelo deve ser



ajustado para obter estimadores mais acurados, dedica-se também a atender pressupostos da análise de regressão. Aqui surge um dos motivos que Sanquetta et al. (2015) mencionam como uma vantagem da aplicação das técnicas de aprendizagem de máquina, a não necessidade de atender condicionantes referentes a testes de hipóteses.

#### 4.5.1.1 Predição da base de dados 98/2012 para validação

Conforme o Algoritmo 17, foi realizada a predição da base de dados para validação. Vale ressaltar que nesse caso a base não apresenta todas as classes de idades, dessa forma optou-se por apresentar uma versão resumida do algoritmo, suprimindo as classes que não estavam presentes.

ALGORITMO 17 – REPRESENTAÇÃO DA PROGRAMAÇÃO DE EQUAÇÕES NÃO-LINEARES PARA ESTIMAR A BIOMASSA DA BASE DE DADOS DE VALIDAÇÃO EM LINGUAGEM DE PROGRAMAÇÃO JULIA.

```
julia> # Definição dos parâmetros
julia> J = validacao[1.<validacao.i.<=2,: ]
julia> MI = validacao[2.<validacao.i.<=4,: ]
julia> MA = validacao[4.<validacao.i.<=7,: ]
julia> # Equações
julia> function Equacao(dados)
    Ymj=(0.205422.*((MJ.d.^2).*MJ.h).^0.459328).+(0.17493*((MJ.d.^2).*MJ.h).
    ^0.506569)
    Yj=(0.033033.*((J.d.^2).*J.h).^0.827421).+(0.043660*((J.d.^2).*J.h).^0.8
    67317)
    Ymi=(0.003976.*((MI.d.^2).*MI.h).^1.056877).+(0.051761*((MI.d.^2).*MI.h)
    .^0.863215)
    Yma=(0.008456.*((MA.d.^2).*MA.h).^0.931885).+(0.087331*(MA.d.^2.146802).
    *(MA.h.^0.353867))
    Ym=(0.000059.*((M.d.^2).*M.h).^1.513338).+(0.146367*(M.d.^2.013462).*(M.
    h.^0.350896))
    yhat = vcat(Yj, Ymi, Yma)
    end
julia> # Obter estimativa
julia> Equacao(validacao)
```

FONTE: O autor (2020).



Os resultados de viés, precisão e acurácia, obtidos para os dados de validação, estão apresentados na Tabela 21. Observa-se que os modelos de regressão *WNSUR* foram muito bem para estimar os dados para treinamento (670/2014), contudo seu desempenho ao estimar a base de dados de validação (98/2012) ficou aquém, evidenciando menor generalização quando comparado às técnicas de aprendizagem de máquina.

TABELA 21 – RESULTADOS DE VIÉS, PRECISÃO E ACURÁCIA PARA OS ESTIMADORES DA VARIÁVEL BIOMASSA DE ÁRVORES DE ACÁCIA-NEGRA UTILIZANDO MODELOS DE REGRESSÃO, PARA A BASE DE DADOS DE **VALIDAÇÃO 98/2012**.

Estimador	$\bar{e}$	$\bar{e}_{\%}$	$S_e$	$S_{e\%}$	$m_x$	$m_{x\%}$
Equações <i>WNSUR</i>	19,8041	26,45	52,8864	70,65	56,4728	75,44

FONTE: O autor (2020).

#### 4.6 MODELAGEM POR MEIO DE APRENDIZAGEM DE MÁQUINA

Foi verificado se os resultados obtidos com a modelagem da biomassa por meio da aprendizagem de máquina são ao menos semelhantes aos resultados com o estimador *WNSUR*, denotado como o principal estimador para regressão de biomassa.

Técnicas de aprendizagem de máquina são procedimentos para classificação de elementos, predição de valores e extração de regras. Por apresentarem uma boa generalidade, podem ser usadas em várias áreas do conhecimento e revelam ótimos resultados, muitas vezes superiores aos obtidos com técnicas convencionais, como pode ser observado em (BINOTI, 2012; WOJCIECHOWSKI, 2015; SANQUETTA et al., 2015; MONTAÑO et al., 2017). Esses pontos são concordantes com os resultados obtidos nesta pesquisa.

Montaño et al. (2017) descreveram que as técnicas de aprendizagem de máquina podem substituir os modelos alométricos que são classicamente utilizados na área de mensuração florestal, o que evidencia que as técnicas de aprendizagem de máquina devem ser consideradas pelos Engenheiros Florestais como uma alternativa viável e segura à regressão. Ainda, os autores consolidam que a regressão, apesar de simples no seu manuseio e ajuste, pode não apresentar a melhor solução para mensuração florestal.

Uma desvantagem, de algumas técnicas de aprendizagem de máquina, é a falta de legibilidade dos modelos, ou seja, em muitos casos, não se consegue explicar

como é a representação dos modelos que foram criados para estimar as variáveis dependentes.

Notavelmente, os Engenheiros Florestais dominam o campo da regressão, a qual é aplicada por mais de meio século na Ciência Florestal brasileira e está presente nos currículos da graduação. Os modelos de aprendizagem de máquina têm conquistado alguns pesquisadores e também estudantes, especialmente aqueles que apreciam a computação, o que é notório pelo aumento do número de pesquisas produzidas por diferentes pessoas na Engenharia Florestal. As aplicações de aprendizagem de máquina estão além das academias, uma vez que algumas empresas adotaram essa alternativa de modelagem. Dados esses fatos, alguns cursos de Engenharia Florestal possuem disciplina a respeito, como é o caso da Disciplina Métodos de Inteligência Artificial aplicados na graduação em Engenharia Florestal ofertada na Universidade Federal do Paraná.

Cada método de aprendizagem de máquina possui suas peculiaridades e considerando que os métodos são complexos, é evidente que sua interpretação é mais difícil que os resultados da análise de regressão. A literatura auxilia a estabelecer uma base para os parâmetros de cada técnica de aprendizagem de máquina, de modo a indicar os hiperparâmetros a serem otimizados. Ao comparar aprendizagem de máquina com a aplicação de regressão, deve-se salientar a similaridade na exaustão em encontrar os melhores resultados para cada método, ao se fazer uma pesquisa visando criar equações de regressão. Sugere-se aos Engenheiros Florestais considerar técnicas de aprendizagem de máquina, visto que cada vez mais técnicas refinadas estão sendo criadas e gerando bons resultados, muitas vezes superiores aos apresentados pelo método de regressão.

Atualmente, a aprendizagem de máquina é onipresente e os modeladores de diferentes domínios da ciência estão tentando adotá-lo, como nesse trabalho, com aplicações na Ciência Florestal. Ao mesmo tempo da conquista de bons resultados com a técnica, a escala e a complexidade das técnicas de aprendizagem de máquina, que são cada vez maiores, torna o seu entendimento um desafio para os usuários. Talvez, o aumento do número de aplicações dessa forma de modelagem irá revelar a experiência necessária para compreensões mais detalhadas das técnicas de aprendizagem de máquina. Fato é que o modelador necessita de experiência sobre as técnicas de aprendizagem de máquina, muito mais do que quando se trabalha com regressão não linear, por exemplo.

A adoção da aprendizagem de máquina tem mudado a maneira de pensar no processamento da modelagem. Isso promoveu uma nova dimensão na modelagem, ao invés de adaptar os modelos a forma das relações das variáveis dependentes e independentes em uma abordagem passo a passo (método paramétrico), o processo de construção da modelagem passou a ser baseada na aprendizagem das relações (método não paramétrico).

A vantagem inerente da aprendizagem é a capacidade de uma técnica aprender os padrões intrínsecos e ocultos nos dados e reagir a novos cenários. Isso ficou claro ao evidenciar as estatísticas obtidas no conjunto de dados de validação, em que as relações alométricas estavam no mesmo domínio, porém conforme exposto na seção análise exploratória dos dados, notou-se que as potências eram diferentes. Nesse sentido, recomenda-se que dados de validação, para avaliação da modelagem de biomassa, não devem ser tomados de madeira aleatória em um conjunto de dados, mas sim, serem compostos por povoamento ou pelo menos por classe de idades dos povoamentos, ou seja, por estratos de povoamentos. O motivo é manter as diferentes relações alométricas que ocorrem nessas condições, pois elas não são de efeito fixo. Ao tomar amostras aleatórias em base de dados, provavelmente essas relações não serão mantidas. Aliás, nesse caso, a amostra é dependente do conjunto de dados e a avaliação de generalização fica comprometida.

Péllico Netto e Behling (2019) descrevem que o modelo alométrico de biomassa pode ser pensado como a distribuição das variáveis fator de forma e densidade do componente da biomassa sobre as variáveis independentes diâmetro à altura do peito e altura, conforme modelo proposto pelos autores. Sabe-se que o fator de forma varia de povoamento para povoamento nas florestas de acácia-negra (MORA; HIGA, 2000), assim como a densidade da madeira (HILLIG et al., 2004; SCHNEIDER et al., 2005). A estratégia em utilizar uma base de dados de validação coletada em outro cenário (base de dados 98/2012), porém com as mesmas características de avaliação das variáveis da base de dados de treinamento (670/2014), foi baseada justamente na hipótese que o fator de forma e a densidade variam nos povoamentos, conseqüentemente, as relações alométricas variam entre os povoamentos e os modelos de aprendizagem de máquina deveriam captar essas informações latentes no conjunto de dados. O que pode ser evidenciado por meio da Figura 15, ao avaliar a relação da biomassa com a variável diâmetro à altura do peito

e altura para as duas bases de dados. Esse ponto desejável, de obter maior generalidade nas estimativas, foi alcançado.

#### 4.7 LINGUAGEM DE PROGRAMAÇÃO *JULIA* NA APRENDIZAGEM DE MÁQUINA

Unindo paradigmas, que muitas vezes estavam distantes, a linguagem de programação *Julia* combina conhecimentos das áreas da ciência da computação para criar uma nova abordagem para informática: a linguagem de programação *Julia* foi projetada pensando na usabilidade e processamento computacional. A experiência obtida em relação a esses dois aspectos, na visão de um Engenheiro Florestal, foi:

##### 4.7.1 Aspecto de Usabilidade

No início da utilização da linguagem de programação, no ano de 2018, várias dificuldades foram observadas, que muitas vezes eram desestimulantes. Alguns pontos da experiência em utilizar ou questionamentos são destacados na sequência:

- *Por que usar a linguagem de programação Julia? Por que não utilizar R, SAS, WEKA ou outra?*
- *Julia* ainda era desconhecida, o que gerava insegurança em sua utilização. Havia uma pressão para utilização de linguagens mais populares e bem desenvolvidas na ciência.
- Havia poucos materiais didáticos. A maioria deles surgiram justamente a partir do ano de 2018. Além do material estar escrito em outros idiomas e a necessidade de importação de livros. As vídeo-aulas disponibilizadas, principalmente por universidades dos Estados Unidos e Inglaterra, também foi uma alternativa importante para aprender a linguagem.
- A versão da linguagem de programação *Julia* disponível na época ainda era instável, até surgir a versão 1.0.
- Aprender a linguagem, principalmente em relação aos primeiros passos, não foi uma tarefa trivial, principalmente devido à pouca experiência de lidar com linguagens de programação, aliás isso não é específico

somente para a linguagem de programação *Julia*, mas também para outras.

- Após um ano de pesquisa sobre a linguagem, trabalhar com *Julia* foi satisfatório, porque os códigos se demonstravam simples dada algumas experiências em outras linguagens.
- Atualmente, já existem muitos materiais didáticos (mesmo que no exterior) e aprender a linguagem de programação *Julia* passou a ser uma tarefa estimulante, principalmente ao evidenciar os resultados obtidos e também saber que universidades e instituições importantes a usam.
- Após esses dois anos de pesquisa (2018-2019), na opinião do autor, a linguagem de programação *Julia* atendeu o aspecto de ser *fácil*. Aliás, isso pode ser evidenciado por meio dos códigos produzidos para realização da modelagem.

#### 4.7.2 Aspecto de desempenho computacional

Neste trabalho, não foram realizadas análises de *benchmark*, o que dificulta a análise desse aspecto cientificamente.

Na literatura, algumas pesquisas foram conduzidas nesse sentido, tais como:

- Aruba e Fernández-Villaverde (2014) descreve que *Julia v.0.2.1* entrega uma performance excelente, tendo uma velocidade de execução entre 2,64 e 2,70 vezes mais lenta que a melhor linguagem, *C v14.0.2*, é ressaltado que a linguagem é pouco mais rápida que a *Java* (JDK8u5) e aproximadamente 4 vezes mais rápida que *Matlab* (2014a).
- Bezanson et al. (2018) comparou linguagens disponíveis no *Programming Language Benchmark Game (PLBG)*. Conforme a metodologia proposta por *PLBG*, os autores mediram o tamanho dos programas removendo comentários e caracteres em branco duplicados, executando a compactação *GZip*. O tamanho combinado de todos os *benchmarks* foi: 6 KB para *Julia v.0.6.2*, 7,4 KB para *JavaScript*, 8,3 KB para *Python 3.5.3* e 14,2 KB para *C*. Comparando a performance de três linguagens (*Julia*, *Python* e *JavaScript*) com os resultados normalizados para o tempo de

processamento para linguagem C. Os resultados demonstraram que *Julia* supera consistentemente *Python* e *JavaScript*, a linguagem está na maioria das vezes dentro de 2 vezes a velocidade de C. Com esses pontos e outros não mencionados os autores afirmam:

“While our observations are encouraging, Julia is still a young language. More experience is needed to draw definitive conclusions as most programs are small and written by domain experts. How the approach we describe here will scale to large (multi-million lines long) programs and to domains outside of scientific computing is a question we hope to answer in future work.”  
(BEZANSON et al., 2018)

A aplicação de aprendizagem de máquina na base de dados 670/2014 já foi realizada em outras linguagens de programação. Com a experiência obtida em linguagem de programação *Julia*, assim como evidenciado na literatura, destaca-se que de fato, a linguagem de programação *Julia* pode ser usada sem comprometer aspectos de desempenho computacional.

#### 4.7.3 Aspectos gerais da linguagem de programação *Julia*

Bezanson et al. (2017) descrevem que a linguagem de programação *Julia* e seu *design* envolvem uma especialização e abstração. Os autores ainda comentam que introduziram em seu *design* a especialização por permitir tratamento personalizado. Ainda, a expedição múltipla (*multiple dispatch*) por ser uma técnica de ciência computacional que escolhe a técnica ou algoritmo apropriado para cada circunstância.

Ao decorrer de toda a construção desse trabalho notou-se que ainda há pontos que ainda devem ser acrescidos para que se tenha uma linguagem de programação para atenderas demandas exigidas por pesquisadores. Notou-se problemas, como por exemplo, a incompatibilidade de certas combinações de *packages*, que causaram problemas na resolução das tarefas. Contudo, deve-se lembrar que a linguagem é nova e vem sendo atualizada recorrentemente, onde, esses problemas tendem a ser resolvidos em futuras atualizações. Principalmente, porque a linguagem tem sido aplicada em múltiplas áreas de pesquisas.

Aplicações da linguagem *Julia* é notória em vários projetos de pesquisas, sendo as principais no ano de 2019, conforme Julia Language (2020):

- *Addition of Baseline Models to Model Zoo*, desenvolvido por SHREYAS KOWSHIK.
- *De-Bruijn Graph Constructor Package for De-novo Genome Assembly*, desenvolvido por Arda Akdemir.
- *Differentiable Duckietown*, desenvolvido por Tejan Karmali.
- *Differentiable Ray Tracer in Julia*, desenvolvido por Avik Pal.
- *Differentiable Tensor Networks*, desenvolvido por Andreas Peter.
- *Enriching Model Zoo with Deep Learning Models*, desenvolvido por Manjunath Bhat.
- *GraphBLAS Implementation*, desenvolvido por Abhinav Mehndiratta.
- *Named Entity Recognition and Part-of-Speech APIs and models*, desenvolvido por Ayushk4.
- *Native Julia ODE, SDE, DAE, DDE, and (S)PDE Solvers*, desenvolvido por Deepesh Thakur.
- *Performance Enhancements and General Fixes*, desenvolvido por Kanav Gupta.
- *Query.jl to SQL translation*, desenvolvido por Brandom Taylor.
- *Rejuvenating Documenter*, desenvolvido por Morten Piibeleht.
- *Towards Better Images.jl Ecosystem*, desenvolvido por Johnny Chen.

Ainda, em relação ao desenvolvimento de outras atividades de pesquisa e referentes a aprendizagem de máquina em *Julia*, destacam-se:

- Chen et al. (2016), que propuseram *MXNet*, a multi-language machine learning.
- Yuret (2016), que desenvolveu *Knet: beginning deep learning with 100 lines of Julia*.
- Innes (2018), o qual propôs *Flux: Elegant machine learning with Julia*.
- Blaom et al. (2019) desenvolveram *MLJ, an open-source machine learning toolbox written in pure Julia*. O *MLJ* é um projeto ambicioso e os desenvolvedores solicitam colaborações para o desenvolvimento do projeto, conforme pode ser avaliado em: <https://github.com/alan-turing-institute/MLJ.jl>. A proposta é fornecer uma interface uniforme para interagir com modelos de aprendizado supervisionado e não supervisionado em diferentes *packages* da linguagem de programação *Julia*. O desenvolvimento do projeto iniciou no *The*



Alan Turing Institute (<https://www.turing.ac.uk/research/projects/machine-learning-julia>), em dezembro de 2018.

Destaca-se que a aprendizagem de máquina requer excelente suporte de *hardware*, bons números de repetições na base de dados, entre outras características, como por exemplo, a aplicação do paralelismo. Onde linguagens de programação de uso geral se esforçam para fornecer esses recursos, o *TensorFlow.jl* da linguagem de programação *Julia* pode lidar com eles sem problemas, embora isso não foi temática desta pesquisa. Em uma palestra proferida por Jonathan Malmaud (pesquisador do *Instituto de Tecnologia de Massachusetts*), disponibilizada em *JuliaCon* 2017, foi demonstrado como as interfaces da linguagem de programação *Julia* com as estruturas populares de aprendizagem de máquina o tornam fácil de usar, em um exemplo com *TensorFlow*.

A base de desenvolvedores da linguagem de programação *Julia* tem aumentado nos últimos meses e as aplicações já são realizadas em um amplo espectro de domínios da ciência, uma lista completa é fornecida na página *Julia Observer* (<https://juliaobserver.com/>). No momento, *Julia* é usada em vários domínios, desde a otimização da produção de leite em fazendas até a supercomputação paralela da astronomia, aliás, pode-se agora descrever, desde a aplicação de uma simples estatística descritiva de diâmetros de árvores. Em uma entrevista para a *Forbes* (D'CUNHA, 2017), Viral Shah, CEO da *Julia Computing*, relata:

*“Amazon, Apple, Disney, Facebook, Ford, Google, Grindr, IBM, Microsoft, NASA, Oracle and Uber are other Julia users, partners and organizations hiring Julia programmers,” says Shah, CEO of Julia Computing. In June, the two-year-old startup raised \$4.6 million seed funding from U.S.-based investment firms General Catalyst and Founder Collective. “Today, Julia has over 1 million downloads, with 161% annual growth”.*

Notavelmente, a linguagem de programação *Julia* chama atenção de algumas das organizações de destaque mundial. Com base na experiência obtida no desenvolvimento deste trabalho, a visão do autor, é que a linguagem de programação *Julia* poderá emergir como uma excelente opção para implementar soluções de aprendizagem de máquina e também uma alternativa para outros tipos de tarefas de processamento de dados na Ciência Florestal.



Um Engenheiro Florestal e área afins, mesmo que experiente e com domínio de outras linguagens de programação, não deve se abster de experimentar *Julia*, porque isso lhe dá outra opção e pode ajudá-lo a descobrir novas maneiras de trabalhar e resolver problemas de modelagem. Entendeu-se também que todas as linguagens de programação são interessantes, úteis, todas com aspectos positivos e negativos e que a linguagem de programação *Julia* é uma alternativa que pode ser aplicada para o processamento de modelagens na Ciência Florestal.

#### 4.7.4 Linguagem de programação *Julia* na área florestal

Conforme mencionado anteriormente, a linguagem de programação *Julia* ainda é recente, e portanto, pouca conhecida e usada na área da ciência florestal. Entretanto tem-se início ao uso da mesma com Engenheiros Florestais como nos seguintes exemplos:

- *Estimating climate service value in forestry: The case of climate information on drought for maritime pine in Southwestern France*, proposto por Cauria e Lobianco (2019).
- *Numerical analysis of ecology-economic model for forest fire fighting in Baikal region*, proposto por Sukhodolov et al. (2019).

Antonello Lobianco, que trabalha na área florestal também desenvolveu um livro intitulado: *Julia Quick Syntax Reference: A Pocket Guide for Data Science Programming* (LOBIANCO, 2019). Na parte de *packages*, é possível encontrar o arquivo *ForestBiometrics.jl*, desenvolvido para geração de gráficos específicos para biometria florestal bem como modelagens específicas para esse campo.

Em termos comerciais, é possível observar a Iniciativa do Custo Social do Carbono (*The Social Cost of Carbon Initiative*), criada pela associação *Resources for the Future (RFF)*, para melhorar a ciência por trás das estimativas do custo social do carbono. As análises são realizadas usando o *package Mimi.jl*, da linguagem de programação *Julia*, focado em análise de modelos de associação integrada (RFF, 2020).

#### 4.8 AVALIAÇÃO DAS HIPÓTESES FORMULADAS NA PESQUISA

Um bom estimador deve ser imparcial. Para que tal sentença seja assegurada, uma distribuição de estimativa deve resultar em viés geral de zero, da mesma forma o estimador deve ser preciso, ou seja, deve apresentar pequena variação e por fim o mesmo deverá ser acurado, sendo que isso significa que seus resultados estão próximos dos valores reais (KOTZ; JOHNSON, 1989).

Conforme mencionado anteriormente, muitas vezes, o principal problema está em um modelo ficar demasiadamente acurado ao ser ajustado, contudo, sua capacidade de prever outros dados será reduzida. Esse é o caso apresentado nos modelos *WNSUR*, visto que se tem um melhor resultado de viés, precisão e acurácia para esse procedimento, em comparação ao melhor estimador de aprendizagem de máquina. Nesse caso, o aprendizado teve resultados melhores apenas para a estatística viés. Os modelos apresentados por Behling et al. (2018) tiveram resultados 3,66% maiores que a técnica de floresta aleatória. Ao analisar-se a precisão e a acurácia apresentadas pela floresta aleatória foram 6,46% e 5,41% superiores aos valores apresentados com os modelos não lineares.

Em contrapartida, ao se observar os dados de validação observa-se uma outra percepção, visto que os valores comparativos de viés, precisão e acurácia apresentam resultados de 14,03%, 55,31% e 55,70% maiores para os modelos *WNSUR*, ao se referir a viés, precisão e acurácia, respectivamente. Sendo assim, é evidenciada a menor capacidade de generalização dos modelos de regressão quando comparado as técnicas de aprendizagem de máquina.

Como o maior objetivo de um estimador não é ser somente bom para estimar os dados utilizados na regressão, mas também outras bases de dados, define-se que as florestas aleatórias tiveram um melhor resultado, visto que seus resultados não foram muito inferiores aos apresentados pelos modelos *WNSUR* no treinamento, entretanto, seus resultados foram amplamente superiores ao comparar-se a validação.

Dessa forma, a primeira hipótese delineada neste trabalho: *“Técnicas de aprendizagem de máquina estimam a biomassa de maneira mais acurada que os modelos de regressão, porque são mais flexíveis e possuem maior capacidade de extrair e generalizar as informações das bases de dados quando comparados com os modelos alométricos”*, foi confirmada conforme resultados já apresentados.

Além da preocupação com a modelagem, o Engenheiro Florestal necessita implementá-la computacionalmente, caso contrário, levaria muito tempo ou seria

inviável. Assim, a modelagem utilizando uma linguagem de programação é fundamental devido a quantidade e diversidade de dados.

Linguagens de programação como *Fortran*, *C*, *Java*, *Python* e outras são utilizadas, e também linguagens especializadas voltadas para programação matemática/estatística, como o *Mathematica*, *Lisp*, *Prolog*, *Matlab*, *Octave*, *R*, *Maple* e *SAS*, podem ser utilizadas no processamento da modelagem. O processo normalmente envolve a análise do problema disponível, escrevendo códigos, analisando o comportamento e a saída, iterando e melhorando o modelo. No final da modelagem, deseja uma solução apropriada para o problema em questão e também espera que os códigos produzidos não sejam longos, que se tenha bom processamento de dados e que os gráficos produzidos tenham alta qualidade.

Conforme apresentado neste trabalho, essas características foram alcançadas utilizando a linguagem de programação *Julia*. Dessa forma, a segunda hipótese delineada neste trabalho: *A linguagem de programação Julia é apropriada para ser utilizada na modelagem de variáveis dendrométricas realizada por Engenheiros Florestais, pois:*

- possui diversas funções matemáticas e estatísticas,
- alta precisão para valores numéricos,
- facilidade para integrar com bibliotecas de outras linguagens,
- os códigos possuem mecanismos para facilitar o entendimento,
- a ideia do programador é expressa utilizando poucas linhas de código, e
- é eficiente em termos de tempo de processamento dos dados,

Tais afirmações foram verificadas conforme resultados já apresentados.

## 5 CONCLUSÕES

A melhor técnica de aprendizagem de máquina para modelar a biomassa de árvores de acácia negra é floresta aleatória com 100 árvores e 1 amostras por folha (*min\_sample\_init*) e classificação *out-of-bag* (*oob\_score*) *true*. Essa técnica é mais apropriada para estimar a biomassa de árvores para novos dados, quando comparado com as equações de regressão ajustadas por *WNSUR*.

A linguagem de programação *Julia* tem potencial para despontar como uma linguagem de programação competitiva e é uma alternativa para problemas que requerem modelagem na Ciência Florestal com exigência computacional. Os códigos são simplificados, seu processamento é de alta performance e a integração com várias bibliotecas é realizada de maneira fácil.

### 5.1 RECOMENDAÇÕES PARA TRABALHOS FUTUROS

Materiais específicos da área florestal utilizando linguagem de programação *Julia* são escassos. Sugere-se a elaboração de materiais a respeito de estatística, manejo florestal, inventário florestal, tecnologia da madeira, ecologia, silvicultura e economia florestal utilizando tal linguagem.

Ao remeter-se as técnicas de aprendizagem de máquina, vale-se destacar que não se tem apenas o método de Predição. Dessa forma, buscar outros enfoques para a área florestal é importante e deve ser feito utilizando-se de outros métodos como: Agrupamento, Associação e Classificação.

## REFERÊNCIAS

- ALIMJAN, G.; SUN, T.; JUMAHUMAN, H.; GUAN, Y.; ZHOU, W.; SUN, H. A hybrid classification approach based on support vector machine and k-nearest neighbor for remote sensing data. **International Journal of Pattern Recognition and Artificial Intelligence**, v. 31, n. 10, p. 1750034. 2017.
- ASSIS, M. R.; TRUGILHO, P.F.; ROSADO, S. C. S.; PROTÁSIO, T. P.; GOULART, S. L. Modelagem da biomassa e do estoque de carbono em plantas jovens de *Eucalyptus*. **Scientia Forestalis**, v. 43, n. 105, p. 225-233. 2015.
- AZEVEDO, T. L.; MELLO, A. A. D.; FERREIRA, R. A.; SANQUETTA, C. R.; NAKAJIMA, N. Y. Equações hipsométricas e volumétricas para um povoamento de *Eucalyptus* sp. localizado na FLONA do Ibura, Sergipe. **Revista Brasileira de Ciências Agrárias (Agrária)**, p. 106-112, 2011.
- AZIZI, Z. Estimating Crown Biomass of Oak Trees Using Terrestrial Photogrammetry in Zagros Forests. In *Advances in Remote Sensing and Geo Informatics Applications*, **Springer**, p. 167-169. 2019.
- BALBAERT, I. Getting started with Julia Programming Language. 214 p., **Packt Publishing**, 2015.
- BALBAERT, I.; SALCEANU, A. Julia 1.0 Programming Complete reference Guide. **Packt Publishing**, 466 p., 2019.
- BALBAERT, I. Julia 1.0 Programming. **Packt Publishing**, 196 p., 2018.
- BARNES, B.V.; ZAK, D. R.; DENTON, S. R.; SPURR, S. H. Forest ecology. 4. ed. New York: **John Wiley & Sons Inc.**, 704p. 1998
- BARNARD, D. M.; GERMINO, M. J.; PILLIOD, D. S.; ARKLE, R. S.; APPLESTEIN, C.; DAVIDSON, B. E.; FISK, M. R. Cannot see the random forest for the decision trees: selecting predictive models for restoration ecology. **Restoration Ecology**. v. 27, n. 5, p. 1053-1065, 2019.
- BASKERVILLE, G. L. Use of logarithmic regression in the estimation of plant biomass. **Canadian Journal of Forest Research**, v. 2, n. 1, p. 49-53. 1972.
- BEHLING, A. Produção de biomassa e o acúmulo de carbono em povoamentos de acácia negra em função de variáveis bioclimáticas. Tese de obtenção de título de mestre em Engenharia Florestal, no curso de Pós-Graduação em Engenharia Florestal da Universidade Federal do Paraná, p. 1-161. Janeiro, 2014.
- BEHLING, A.; SANQUETTA, C. R.; DALLA CORTE, A. P.; NETTO, S. P.; RODRIGUES, A. L.; CARON, B. O.; Simon, A. A. Tracking leaf area index and coefficient of light extinction over the harvesting cycle of black wattle. **Journal of forestry research**, v.27, n. 6, p. 1211-1217, 2016.

BEHLING, A.; PÉLLICO NETTO, S.; SANQUETTA, C. R.; DALLA CORTE, A. P.; AFFLECK, D. L. R.; RODRIGUES, A. L.; BEHLING, M. Critical analyses when modeling tree biomass to ensure additivity of its componentes. **Anais da Academia Brasileira de Ciência**, v. 90, n. 2, p. 1759-774, 2018.

BEHLING, A.; PÉLLICO NETTO, S.; SANQUETTA, C. R.; DALLA CORTE, A. P.; SIMON, A. A.; RODRIGUES, A. L.; CARON, B. O. Additive and Non-additive Biomass Equations for Black Wattle. **Floresta e Ambiente**, v. 26, n. 4. 10p. 2019.

BEHLING, M. Nutrição, partição de biomassa e crescimento de povoamentos de Teca em Tangará da Serra-MT. Universidade Federal de Viçosa, 175p., 2009.

BELLISARIO, K. M.; BROADHEAD, T.; SAVAGE, D.; ZHAO, Z.; OMRANI, H.; ZHANG, S.; PIJANOWSKI, B. C. Contributions of MIR to soundscape ecology. Part 3: Tagging and classifying audio features using a multi-labeling k-nearest neighbor approach. **Ecological Informatics**, v. 51, p. 103-111. 2019.

BEZANSON, J.; CHEN, J.; CHUNG, B.; KARPINSKI, S.; SHAH, V. B.; VITEK, J.; ZOUBRITZKY, L. Julia: dynamism and performance reconciled by design. **Proceedings of the ACM on Programming Languages**, 2(OOPSLA), 120:1-23p. 2018.

BEZANSON, J.; EDELMAN, A.; KARPINSKI, S.; SHAH, V. B. Julia: A Fresh Approach to Numerical Computing. **SIAM Review**, v. 59, n. 1, p. 65-98, 2017.

BEZANSON, J., KARPINSKI, S., SHAH, V., EDELMAN, A., Why We Created Julia, 14 de Fevereiro de 2012, Disponível em: <https://julialang.org/blog/2012/02/why-we-created-julia> Acesso em: 26/12/2019.

BINOTI, D. H. B.; BINOTI, M. L. M. D. S.; LEITE, H. G.; ANDRADE, A. V.; NOGUEIRA, G. S.; ROMARCO, M. L.; PITANGUI, C. G. Support vector machine to estimate volume of Eucalypt trees. **Revista Árvore**, v. 40, n. 4, p. 689-693. 2016.

BINOTI, M. L. M. S. Emprego de redes neurais artificiais em mensuração e manejo florestal. Tese de Mestrado - Universidade Federal de Viçosa, Viçosa, MG, 137p. 2012.

BINOTI, M. L. M. D. S.; BINOTI, D. H. B.; LEITE, H. G.; GARCIA, S. L. R.; FERREIRA, M. Z.; RODE, R.; SILVA, A. A. L. D. Redes neurais artificiais para estimação do volume de árvores. **Revista Árvore**, v. 38, n. 2, p. 283-288, 2014.

BINOTI, M. L. D. S.; BREDI BINOTI, D. H.; GARCIA LEITE, H. Aplicação de redes neurais artificiais para estimação da altura de povoamentos equiâneos de eucalipto. **Revista Árvore**, v. 37, n.4, p. 639-645, 2013.

BINOTI, M. L. M. S.; LEITE, H. G.; BINOTI, D. H. B.; GLERIANI, J. M. Prognose em nível de povoamento de clones de eucalipto empregando redes neurais artificiais. **Revista Cerne**, v. 21, n. 1, p. 97-105, 2015.

BLAOM, A.; ARENAS, D.; KIRALY, F.; SIMILLIDES, Y.; VOLLMER, S. Beyond machine learning pipelines with MJL. 2019 Link: <<https://julialang.org/blog/2019/05/beyond-ml-pipelines-with-mlj>>

BOUREL, M.; SEGURA, A. M. Multiclass classification methods in ecology. **Ecological indicators**, v. 85, p. 1012-1021. 2018.

BREIMAN, L. Random forests. **Machine Learning**, v. 45, n. 1, p. 5–32. 2001.

BROWN, J.H.; GILLOOLY, J.F.; ALLEN, A.P.; SAVAGE, V.M.; WEST, G.B. Toward a metabolic theory of ecology. **Ecology**, v. 85, n. 7, p. 1771-1789, 2004.

CAI, X. H.; CHEN, T.; WANG, R. Y.; FAN, Y. J.; LI, Y.; HU, S. N.; ZHOU, Q. M. Forecasting wildfire disease on tobacco: toward developing a high-accuracy prediction model for disease index using local climate factors and support vector regression. **Theoretical and Applied Climatology**, v. 137, n. 3-4, p. 2139-2149. 2019.

CALDEIRA, M. V. W. Determinação de biomassa e nutrientes em uma floresta ombrófila mista montana em General Carneiro , Paraná. Tese de obtenção de título de mestre em Engenharia Florestal, no curso de Pós-Graduação em Engenharia Florestal da Universidade Federal do Paraná. p. 1-192, 2003.

CAMPOS, B. P. F.; SILVA, G. F.; BINOTI, D. H. B.; MENDONÇA, A. R.; LEITE, H. G. Descrição do perfil do tronco de árvores em plantios de diferentes espécies por meio de redes neurais artificiais. **Pesquisa Florestal Brasileira**, v. 37, n. 90, p. 99-107. 2017.

CARRASQUILLA, J.; MELKO, R. G. Machine Learning Phases of Matter. *Natural Physics*. v. 16, p. 431-434, 2017.

CAVALLARO, G.; RIEDEL, M.; RICHERZHAGEN, M.; BENEDIKTSSON, J. A.; PLAZA, A. On understanding big data impacts in remotely sensed image classification using support vector machine methods. **IEEE journal of selected topics in applied earth observations and remote sensing**, v. 8, n. 10, p. 4634-4646. 2015.

CASTAÑEDA, D. S. **Aplicación de support vector machine al mercadocolombiano**. Doctoral dissertation, Universidad del Rosario. 2019.

CAURLA, S.; LOBIANCO, A. Estimating climate service value in forestry: The case of climate information on drought for maritime pine in Southwestern France. **Climate Service**, 12p. 2019.

CHAVES, J.; ANDALO, C.; BROWN, S.; CAIRNS, M.A.; CHAMBERS, J.Q.; EAMUS, D.; FOLSTER, H.; FROMARD, F.; HIGUCHI, N.; KIRA, T.; LESCURE, J.P.; NELSON, B.W.; OGAWA, H.; PUIG, H.; RIÉRA, B.; YAMAKURA, T. Tree allometry and improved estimation of carbon stocks and balance in tropical forests. **Oecologia**, v. 145, p. 87-99. 2005.



CHEN, J. L.; LIU, H. B.; WU, W.; XIE, D. T. Estimation of monthly solar radiation from measured temperatures using support vector machines-a case study. **Renewable Energy**, v. 36, n. 1, p. 413-420, 2011.

CHEN, T.; LI, M.; LI, Y.; LIN, M.; WANG, N.; WANG, M.; XIAO, T.; XU, B.; ZHANG, C.; ZHANG, Z. A flexible and efficient machine learning library for heterogeneous distributed systems. **arXiv**, v. 1, 6p. 2016.

CHIRICI, G.; MURA, M.; MCINERNEY, D.; PY, N.; TOMPPPO, E. O.; WASER, L. T.; MCROBERTS, R. E. A meta-analysis and review of the literature on the k-Nearest Neighbors technique for forestry applications that use remotely sensed data. **Remote Sensing of Environment**, v. 176, p. 282-294. 2016

CORDEIRO, M. A.; JESUS PEREIRA, N. N.; BINOTI, D. H. B.; SILVA BINOTI, M. L. M.; LEITE, H. G. Estimativa do volume de *Acacia mangium* utilizando técnicas de redes neurais artificiais e máquinas vetor de suporte. **Pesquisa Florestal Brasileira**, v. 35, n. 83, p. 255-261. 2015.

CORTES, C.; VAPNIK, V. Support-vector networks. Machine learning, **Springer**, v. 20, n. 3, p. 273-297, 1995.

COSTA FILHO, S. V. S.; ARCE, J. E.; MONTAÑO, R. A. N. R.; PELISSARI, A. L. Configuração de algoritmos de aprendizado de máquina na modelagem florestal: um estudo de caso na modelagem da relação hipsométrica. **Ciência Florestal**, v. 29, n. 4, p. 1501-1515. 2019.

CUNHA, E. G. S. Quantificação de biomassa em floresta estacional semidecidual por meio de redes neurais artificiais. Dissertação apresentada à Universidade Federal dos Vales do Jequitinhonha e Mucuri, como parte das exigências do Programa de Pós-Graduação em Ciência Florestal. p. 2-76, 2015.

CYBENKO, G. Approximation by super positions of sigmoidal function. **Mathematics of Control, Signals and Systems**, 2 ed., 314p. 1989.

D'CUNHA, S. D. **How a new programming language created by four scientists now used by the world's biggest companies**. Forbes. 2017. Entrevista. Link: <<https://www.forbes.com/sites/suparnadutt/2017/09/20/this-startup-created-a-new-programming-language-now-used-by-the-worlds-biggest-companies/#227cb3837de2>>

DAI, Q. Y.; ZHANG, C. P.; WU, H. Research of decision tree classification algorithm in data mining. **Internacional Journal of Database Theory and Application**, v.9, n. 5, p. 1-8, 2016.

DA SILVA, J. A. A. Potencialidades de florestas energéticas de *Eucalyptus* no Pólo Gesseiro do Araripe-Pernambuco. **Anais da Academia Pernambucana de Ciência Agrônômica**, v. 5, p. 301-319. 2013.



DA SILVA, R. C.; RIBEIR, R. B. D. S.; XIMENES, L. C. Ajuste de equações volumétricas para diferentes seccionamentos de tora em uma floresta sob manejo florestal comunitário. **Acta Tecnológica**, v. 12, n. 2, p. 85-102. 2018.

DE ABREU, J. C.; SOARES, C. P. B.; LEITE, H. G. Assessing alternatives to estimate the stem volume of a seasonal seim-deciduous forest. **Revista Floresta**, v. 47, n. 4, p. 375-382. 2018.

DEBASTINI, A. B.; DE VASCONCELLOS, B. N.; MARTINS, A. P. M.; DALLA-CORTE, A. P.; SANQUETTA, C. R. Predições geoestatísticas e zonas de incerteza no inventário de povoamentos florestais. **Advances in Forestry Science**, v. 5, n. 1, p. 283-287. 2018.

DEKLERCK, V.; FINCH, K.; GASSON, P.; VAN DEN BULCKE, J.; VAN ACKER, J.; BEECKMAN, H.; ESPINOZA, E. Comparison of species classification models of mass spectrometry data: Kernel Discriminant Analysis vs Random Forest; A case study of Afrormosia (*Pericopsis elata* (Harms) Meeuwen). **Rapid Communications in Mass Spectrometry**, v. 31, n. 19, p. 1582-1588. 2017.

DE MELO, L. C.; SANQUETTA, C. R.; DALLA CORTE, A. P.; HENTZ, Â. M. K. Estimativa de biomassa e carbono total para árvores de caixeta no Paraná. **Pesquisa Florestal Brasileira**, v. 34, n. 77, p. 21-30. 2014.

DE MIRANDA, D. L. C.; JUNIOR, V. B.; GOUVEIA, D. M. Fator de forma e equações de volume para estimativa volumétrica de árvores em plantio de *Eucalyptus urograndis*. **Scientia plena**, v. 11, n. 3, 8 p., 2015.

DEO, R.; SAHIN, M. Application of the Artificial Neural Network model for prediction of monthly standardized precipitation and evapotranspiration index using hydrometeorological parameters and climate indices in eastern Australia. **Atmospheric research**, v. 161-162, p. 65-81, 2015.

DE OLIVEIRA, T. W. G.; LIMA, E. M.; SANQUETTA, C. R.; DALLA CORTE, A. P. Ajuste simultâneo na predição de biomassa aérea em *Araucaria angustifolia* (Bertol.) Kuntze. **BIOFIX Scientific Journal**, v. 3, n. 1, p. 137-144. 2018.

DOLÁCIO, C. J. F.; OLIVEIRA, R. S.; NAKAJIMA, N. Y.; SILVA, D. A.; EBLING, Â. A.; ROCHA, J. E. C. Modelagem do carbono estocado no tronco de árvores de *Eucalyptus* com modelos lineares e rede neural artificial. **Advances in Forestry Science**, v. 6, n. 2, p. 691-697. 2019.

DOS REIS MARTINS, E.; SILVA BINOTI, M. L. M.; LEITE, H. G.; BINOTI, D. H. B.; Dutra, G. C. Configuração de redes neurais artificiais para estimação do afilamento do fuste de árvores de eucalipto. **Revista Brasileira de Ciências Agrárias**, v. 11, n. 1, p. 33-38. 2016.

DOS SANTOS, G. L.; PEREIRA, M. G.; NETO, T. D. A. C.; MENDONÇA, V. M. M.; MENEZES, C. E. G. Ciclagem de nutrientes em diferentes condições topográficas em Floresta Estacional Semidecidual, Pinheiral-RJ. **Ciência Florestal**, v. 29, n. 4, p. 1737-1747. 2019.

DOS SANTOS, K. N. F.; RODE, R.; DE ANDRADE, D. F. C.; DOS SANTOS CORRÊA, K. K.; DE SOUZA LOPES, L. S. Ajuste de equações volumétricas e redes neurais artificiais na estimativa do volume de Tauari na Floresta Nacional do Tapajós. **Revista Agroecossistemas**, v. 10, n. 1, p. 1-17. 2018.

DUDA, R. O.; HART, P.E.; STORK, D. G. Pattern Classification. 2 ed., **Wiley-Interscience**, 2000.

DUDANI, S. A. The distance-weighted k-nearest-neighbor rule. **IEEE Transactions on Systems, Man, and Cybernetics**, n. 4, p. 325-327, 1976.

DURANTE, P.; MARTÍN-ALCÓN, S.; GIL-TENA, A.; ALGEET, N.; TOMÉ, J. L.; RECUERO, L.; OYONARTE, C. Improving Aboveground Forest Biomass Maps: From High-Resolution to National Scale. **Remote Sensing**, v. 11, n. 7, p. 795. 2019.

FAO. State of World's Forests 2011. Rome, 2011, 164 p. Disponível em: <<http://www.fao.org/docrep/013/i2000e/i2000e.pdf>>. Acesso em: 30 março de 2019.

FENG, Y.; ZHANG, W.; SUN, D.; ZHANG, L. Ozone concentration forecast method based on genetic algorithm optimized back propagation neural networks and support vector machine data classification. **Atmospheric Environment**, v. 45, n. 11, p. 1979-1985. 2011.

FERREIRA, J. C.; STÄHELIN, T. S. F.; JESUS, M. S. D.; MUÑIZ, G. I. B. D.; BRAND, M. A.; FREITAS, T. P. Estimativa da oferta de biomassa florestal em povoamentos de *Pinus taeda* L. após intervenções culturais. **Ciência Florestal**, v. 29, n. 3, p. 1459-1468. 2019.

FIGUEIREDO, S. D. M.; & FIGUEIREDO, E. O. Aplicação de dados de ocorrência de inventário florestal na modelagem de distribuição de espécies florestais madeireiras no leste do Acre, Brasil. In Embrapa Acre-Artigo em anais de congresso (ALICE). In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 19., 2019, Santos, SP. **Anais. INPE**, 4p., 2019.

FINLEY, A. O.; DATTA, A.; COOK, B. D.; MORTON, D. C.; ANDERSEN, H. E.; BANERJEE, S. Efficient algorithms for Bayesian nearest neighbor Gaussian processes. **Journal of Computational and Graphical Statistics**, p. 1-14. 2019.

FONSECA, F.; FIGUEIREDO, T. D.; VILELA, Â.; SANTOS, R.; CARVALHO, A. L.; ALMEIDA, E.; NUNES, L. Impacto da substituição de espécies florestais no armazenamento de carbono em áreas de montanha da região mediterrânea. **VIII Congresso Ibérico de Ciências do Solo**. p. 594-597., 2018.

FOX, E. W.; HILL, R. A.; LEIBOWITZ, S. G.; OLSEN, A. R.; THORNBRUGH, D. J.; WEBER, M. H. Assessing the accuracy and stability of variable selection methods for random forest modeling in ecology. **Environmental monitoring and assessment**, v. 189, n. 7, 316p. 2017.

GHOSH, S. M.; BEHERA, M. D. Aboveground biomass estimation using multi-sensor data synergy and machine learning algorithms in a dense tropical forest. **Applied Geography**, v. 96, p. 29-40. 2018.

GLEASON, C. J.; IM, J. Forest biomass estimation from airborne LiDAR data using machine learning approaches. **Remote Sensing of Environment**, v. 125, p. 80-91. 2012.

GOLLAPUDI, S. Practical Machine Learning – Tackle the real-world complexities of modern machine learning with innovative and cutting-edge techniques. **Packt Publishing Ltd.**, 419 p., 2016.

GOLLEY, F.B. MCGINNIS, J. T.; CLEMENTS, R. G.; CHILD, G. I.; DUEVER, M. J. Ciclagem de minerais em um ecossistema de floresta tropical úmida. **São Paulo: USP**. 256p. 1978.

GREENE, W.H. Econometric analysis. **Upper Saddle River: Prentice-Hall**, 1178p. 2008.

GUENTHER, N.; SCHONLAU, M. Support vector machines. **The Stata Journal**, v. 16, n. 4, p. 917-937. 2016.

HAAG, H.P. Ciclagem de nutrientes em florestas tropicais. **Campinas: Fundação Cargil**. 144p. 1985.

HART, E.; SIM, K.; GARDINER, B.; KAMIMURA, K. A hybrid method for feature construction and selection to improve wind-damage prediction in the forestry sector. **In Proceedings of the Genetic and Evolutionary Computation Conference**, p. 1121-1128. 2017.

HASTIE, T.; TIBSHIRANI, R. FRIEDMAN, J. The Elements of Statistical Learning: Data Mining, Inference and Prediction. **Springer, corrected edition**. 2003.

HASTIE, T.; TIBSHIRANI, R. Discriminant adaptive nearest neighbor classification and regression. **In Advances in Neural Information Processing Systems**, p. 409-415. 1996.

HAYKIN, S. Redes neurais: princípios e prática. **Porto Alegre: Bookman**, 2001.

HAYKIN, S.; VEEN, B. V. Signals and systems. **[S.I.]: Wiley**, 1999.

HECHENBICHLER, K.; SCHLIEP, K. Wighted k-nearest-neighbor techniques and ordinal classification. **Munchen: Ludwig-Maximilians-Universitat**, 17p, 2004.

HIGUCHI, N.; CARVALHO Jr., J. A. Biomassa florestal e conteúdo de carbono de espécies arbóreas da Amazônia. In: Emissão x Sequestro de CO<sub>2</sub> - Uma Nova Oportunidade de Negócios para o Brasil. **Companhia Vale do Rio Doce, Rio de Janeiro, RJ, Brasil**. p. 125-153. 1994.

HILLIG, E.; HASELEIN, C. R.; SANTINI, E. J. Estabilidade dimensional de chapas aglomeradas estruturais (flakeboards) fabricadas com madeiras de pinus, eucalipto e acácia-negra. **Scientia Forestalis**, v. 65, p. 80-94. 2004.

HOKKANEN, A. H.; HÄNNINEN, L.; TIUSANEN, J.; PASTELL, M. Predicting sleep and lying time of calves with a support vector machine classifier using accelerometer data. **Applied Animal Behaviour Science**, v. 134, n. 1-2, p. 10-15. 2011.

HOPFIELD, J. Neural networks and physical systems with emergent collective computationalabilities. **Proceedings of the National Academy of Sciences of the United States of America**, v. 79, p. 2554–2558, 1982.

HUANG, Y. L.; WANG, K. L.; CHEN, D. R. Diagnosis of breast tumors with ultrasonic texture analysis using support vector machines. **Neural Computing & Applications**, v. 15, n. 2, p. 164-169, 2006.

HUNTER, M. O.; KELLER, M.; VICTORIAa, D.; MORTON, D. C. Tree height and tropical forest biomass estimation. **Biogeosciences**, v. 10, n. 12, p. 8385-8399. 2013.

IQBAL, F.; LUCIEER, A.; BARRY, K. Poppy crop capsule volume estimation using UAS remote sensing and random forest regression. **International journal of applied earth observation and geoinformation**, v. 73, p. 362-373. 2018.

INNES, M. Flux: Elegant machine learning with Julia. **The Journal of Open Source Software**. v. 3, n. 25, p. 602, 2018.

JAAFARI, A.; POURGHASEMI, H. R. Factors Influencing Regional-Scale Wildfire Probability in Iran: An Application of Random Forest and Support Vector Machine. In **Spatial Modeling in GIS and R for Earth and Environmental Sciences**. Elsevier, p. 607-619. 2019.

JONES, C. B. Geographical information systems and computer cartography. **Routledge**. 344p. 1997.

JOSHI, A. Julia for Data Science: Explore the world of data Science from scratch with Julia by your side. **Packt Publishing**, p. 1-339, 2016.

JOSHI, A.; LAKHANPAL, R. Learning Julia – Build high-performance applications dor scientific computing. **Packt Publishing**. 316p. 2017.

JULIA LANGUAGE. Link: <<https://summerofcode.withgoogle.com/archive/2019/organizations/4615767377575936/>> Acesso: 09/01/2020.

KAMINSKI, B.; SZUFEL, P. Julia 1.0 Programming Cookbook. **Packt Publishing**, 460 p., 2018.

KANKARE, V.; HOLOPAINEN, M.; VASTARANTA, M.; PUTTONEN, E.; YU, X.; HYYPPÄ, J.; ALHO, P. Individual tree biomass estimation using terrestrial laser

scanning. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 75, p. 64-75. 2013.

KARIMI, F.; SULTANA, S.; BABAKAN, A. S.; SUTHAHARAN, S. An enhanced support vector machine model for urban expansion prediction. **Computers, Environment and Urban Systems**, v. 75, p. 61-75. 2019.

KAUR, H.; SOOD, S. K. Soft-computing-centric framework for wildfire monitoring, prediction and forecasting. **Soft Computing**, p. 1-11. 2019.

KERN, A. N.; ADDISON, P.; OOMMEN, T.; SALAZAR, S. E.; COFFMAN, R. A. Machine learning based predictive modeling of debris flow probability following wildfire in the intermountain Western United States. *Mathematical Geosciences*, v. 49, n. 6, p. 717-735. 2017.

KILHAM, P.; HARTEBRODT, C.; KÄNDLER, G. Generating Tree-Level Harvest Predictions from Forest Inventories with Random Forests. **Forests**, v. 10, n. 20, p. 1-25, 2019.

KISI, O.; ÇİMEN, M. Evapotranspiration modelling using support vector machines. **Hydrology Science Journal**, v. 85, p. 918-928, 2009.

KOHONEN, T. Self-organized formation of topologically correct feature maps. **Biological Cybernetics**, v. 43, p. 59-69, 1982.

KOZLOWSKI, T.T.; PALLARDY, S.G. Physiological of woody. 2. ed. **San Diego: Academic**, 432p. 1996.

KREBS, C. J. Ecological methodology. **Pearson**. 624p. 2007.

KUYAH, S.; DIETZ, J.; MUTHURI, C.; JAMNADASS, R.; MWANGI, P.; COE, R.; NEUFELDT, H.; Allometric equations for estimating biomass in agricultural landscapes: I. Aboveground biomass. **Agriculture, Ecosystems and Environment**, v. 158, p. 216-224, 2014.

KWON, C. Julia Programming for Operations Research. **Independently published**. 260 p., 2019.

LARCHER, W. Ecofisiologia vegetal. **São Carlos: Rima**, p. 525-550, 2006.

LAU, A.; CALDERS, K.; BARTHOLOMEUS, H.; MARTIUS, C.; RAUMONEN, P.; HEROLD, M.; GOODMAN, R. C. Tree Biomass Equations from Terrestrial LiDAR: A Case Study in Guyana. **Forests**, v. 10, n. 6, p. 527. 2019.

LAUWENS, B.; DOWNEY, A. B. Think Julia - How to Think Like a Computer Scientist. **O'Reilly Media**, 298 p., 2019.

LIU, Z.; PENG, C.; WORK, T.; CANDAU, J. N.; DESROCHERS, A.; KNEESHAW, D. Application of machine-learning methods in forest ecology: recent progress and future challenges. **Environmental Reviews**, v. 26, n. 4, p. 339-350. 2018.

LIU, Z.; ZHOU, P.; CHEN, G.; GUO, L. Evaluating a coupled discrete wavelet transform and support vector regression for daily and monthly streamflow forecasting. **Journal of hydrology**, v. 519, p.2822-2831, 2014.

LI, Y.; DONG, W.; YANG, Q.; ZHAO, J.; LIU, L.; FENG, S. An Automatic Impedance Matching Method Based on the Feedforward-Backpropagation Neural Network for a WPT System. **IEEE Transactions on Industrial Electronics**, v. 66, n. 5, p. 3963-3972. 2018.

LOBIANCO, A. Julia Quick Syntax Reference: A Pocket Guide for Data Science Programing. **Apress**, 216p. 2019.

LOPES, L. D. S.; RODE, R.; PAULETTO, D.; BALONEQUE, D. D.; SILVA, A.; DOS SANTOS, K. N. F. Ajuste de modelos de taper e sortimento de toras de mogno africano em sistemas agroflorestais em Belterra, Pará. **Revista Agroecossistemas**, v. 10, n. 1, p. 18-28, 2018.

LORENA, A. C.; JACINTHO, L. F.; SIQUEIRA, M. F.; DE GIOVANNI, R.; LOHMANN, L. G.; DE CARVALHO, A. C.; YAMAMOTO, M. Comparing machine learnin classifiers in potential distribution modelling. **Expert Systeme with Applications**, v. 38, n. 5, p. 5268-5275, 2011.

LOU, I.; XIE, Z.; UNG, W. K. ; MOK, K. M. Integranting support vector regression with particle swarm optimization for numerical modeling for algal blooms of freshwater. In Advanes in Monitoring and Modelling Algal Blooms in Freshwater Reservoirs. **Springer, Dordrecht**, p. 125-141, 2017.

LUNA, A. S.; PAREDES, M. L. L.; DE OLIVEIRA, G. C. G.; CORRÊA, S. M. Prediction of ozone concentration in tropospheric levels using artificial neural networks and support vector machine at Rio de Janeiro, Brazil. **Atmospheric Environment**, v. 98, p. 98-104. 2014.

LYRA, G. B.; ZANETTI, S.S.; SANTOS, A. A. R.; DE SOUZA, J. L.; OLIVEIRA-JÚNIOR, J. F.; LEMES, M. A. M. Estimation of monthly global solar irradiation using the Hargreaves-Samani model and na artificialneural network for the state of Alagoas in northeastern Brazil. **Theoretical and applied climatology**, v. 125, n. 3-4, 743-756, 2016.

MACHADO, S. do A.; FIGUEIREDO FILHO, A. Dendrometria. **Curitiba: Editora UNICENTRO**, 316 p., 2003.

MARQUES DA SILVA, M. L.; BREDÁ BINOTI, D. H.; GLERIANI, J. M.; GARCIA LEITE, H. Ajuste do modelo de Schumacher e Hall e aplicação de redes neurais artificiais para estimar volume de árvores de eucalipto. **Revista Árvore**, v. 33, n. 6, p. 1133-1139, 2009.



MARTINELLI, L. A.; MOREIRA, M. Z.; BROWN, I. F.; VICTORIA, R. L. Incertezas associadas às estimativas de biomassa em florestas tropicais: o exemplo de uma floresta situada no estado de Rondônia. **Anais do Seminário de Emissão x Sequestro de CO<sub>2</sub>: uma nova oportunidade de negócios para o Brasil**. p. 1-221, 1994.

MARTINS, A. P. M.; DEBASTIANI, A. B.; PELISSARI, A. L.; MACHADO, S. D. A.; SANQUETTA, C. R. Estimativa do Afilamento do Fuste de Araucária Utilizando Técnicas de Inteligência Artificial. **Floresta e Ambiente**, v. 24, p. 1-10, 2017.

MARTINS, J. F. C.; DA SILVA, S. A.; COUTINHO, V. M.; ORSO, G. A.; BEHLING, A.; DALLA CORTE, A. P. Carbono nos componentes da biomassa de *Acacia mearnsii* De Wild. **BIOFIX Scientific Journal**, v. 5, n. 1, p. 32-38, 2019.

MALEK, S.; MIGLIETTA, F.; GOBAKKEN, T.; NAESSET, E.; GIANELLE, D.; DALPONTE, M. Prediction of stem diameter and biomass at individual tree crown level with advanced machine learning techniques. **iForest-Biogeosciences and Forestry**, v. 12, n. 3, p. 323. 2019.

MAULIK, U.; CHAKRABORTY, D. Remote Sensing Image Classification: A survey of support-vector-machine-based advanced techniques. **IEEE Geoscience and Remote Sensing Magazine**, v. 5, n. 1, p. 33-52. 2017.

MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. **Bulletin of Mathematical Biophysics**, p. 115-133, 1943.

MCCROBERTS, R. E.; NAESSET, E.; GOBAKKEN, T. Optimizing the k-Nearest Neighbors technique for estimating forest aboveground biomass using airborne laser scanning data. **Remote Sensing of Environment**, v. 163, p. 13-22. 2015.

MIGUEL, E. P.; REZENDE, A. V.; LEAL, F. A.; MATRICARDI, E. A. T.; DO VALE, A. T.; PEREIRA, R. S. Redes neurais artificiais para a modelagem do volume de madeira e biomassa do cerrado com dados de satélite. **Pesquisa Agropecuária Brasileira**, v. 50, n. 9, p. 829-839, 2015.

MIKHAIL, E.; ACKERMAN, F. Observations and Least Squares. **Press of America**, p. 497, 1976.

MITCHELL, T. M. Machine Learning. **McGraw-Hill Science/Engineering/Math**. 420p. 1997.

MOMO TAKOUDJOU, S.; PLOTON, P.; SONKÉ, B.; HACKENBERG, J.; GRIFFON, S.; DE COLIGNY, F.; PÉLISSIER, R. Using terrestrial laser scanning data to estimate large tropical trees biomass and calibrate allometric models: A comparison with traditional destructive approach. **Methods in Ecology and Evolution**, v. 9, n. 4, p. 905-916. 2018.

MOMOLLI, D. R.; SCHUMACHER, M. V. Sustentabilidade de povoamentos de *Acacia mearnsii* De Wild. em diferentes idades: uma revisão da ciclagem de nutrientes.

**Revista Brasileira de Gestao Ambiental e Sustentabilidade**, v. 6, n. 12, p. 263-272. 2019.

MONTAÑO, R. A. N. R. Aplicação de técnicas de aprendizado de máquina na mensuração florestal. Tese como requisito parcial para obtenção do título de Doutor em Engenharia Florestal, no Curso de Pós-Graduação em Engenharia Florestal da Universidade Federal do Paraná, p. 1-86, 2016.

MONTAÑO, R. A. N. R.; SANQUETTA, C. R.; WOJCIECHOWSKI, J.; MATTAR, E.; DALLA CORTE, A. P.; TODT, E. Artificial Intelligence Models to Estimate Biomass of Tropical Forest Trees. **Polibits**, v. 56, p. 29-37, 2017.

MORA, A. L.; HIGA, A. R. Manejo das informacoes de pesquisa florestal: o caso da acacia-negra. In Embrapa Florestas-Artigo em anais de congresso (ALICE). In: **Reunion de la red de informacion para America Latina y el caribe**, v. 6., p. 45-51. 2000.

MORIN, M.; GAUDREAU, J.; BROTHERTON, E.; PARADIS, F.; ROLLAND, A.; WERY, J.; LAVIOLETTE, F. Machine learning-based models of sawmills for better wood allocation planning. **International Journal of Production Economics**, p. 107508. 2019.

NAGAR, S. Introduction to Julia Programming - For Scientists and Engineers. **Independently published**, 282 p., 2017.

NIETO, P. G.; TORRES, J. M.; FERNÁNDEZ, M. A.; GALÁN, C. O. Support vector machines and neural networks used to evaluate paper manufactured using Eucalyptus globulus. **Applied Mathematical Modelling**, v. 36, n. 12, p.6137-6145, 2012.

NIKLAS, K.J. Plant allometry: the scaling of form and process. **University of Chicago Press**, 1994. 412p.

NITOSLAWSKI, S. A.; GALLE, N. J.; VAN DEN BOSCH, C. K.; STEENBERG, J. W. Smarter ecosystems for smarter cities? A review of trends, technologies, and turning points for smart urban forestry. **Sustainable Cities and Society**, 101770. 2019.

NUNES, M. H.; GÖRGENS, E. B. Artificial intelligence procedures for tree taper estimation within a complex vegetation mosaic in Brazil. **PloS one**, v. 11, n. 5, p. 0154738. 2016.

OGUNTOKE, O.; EMORUWA, F. O.; TAIWO, M. A. Assessment of air pollution and health hazard associated with sawmill and municipal waste burning in Abeokuta Metropolis, Nigeria. **Environmental Science and Pollution Research**, p. 1-15. 2019.

ORTIZ-REYES, A. D.; VALDEZ-LAZALDE, J. R.; ÁNGELES-PÉREZ, G.; HÉCTOR, M.; SCHNEIDER, L.; AUGIRRE-SALADO, C. A.; PEDUZZI, A. Transectos de datos LiDAR: una estrategia de muestreo para estimar biomasa aérea en áreas forestales. **Madera y Bosques**, v. 25, n. 3. 2019.



OTA, T.; AHMED, O. S.; MINN, S. T.; KHAI, T. C.; MIZOUE, N.; YOSHIDA, S. Estimating selective logging impacts on aboveground biomass in tropical forests using digital aerial photography obtained before and after a logging event from an unmanned aerial vehicle. **Forest ecology and management**, v. 433, p. 162-169. 2019.

OTERO, V.; VAN DE KERCHOVE, R.; SATYANARAYANA, B.; MARTÍNEZ-ESPINOSA, C.; FISOL, M. A. B.; IBRAHIM, M. R. B.; DAHDOUN-GUEBAS, F. Managing mangrove forests from the sky: Forest inventory using field data and Unmanned Aerial Vehicle (UAV) imagery in the Matang Mangrove Forest Reserve, peninsular Malaysia. **Forest ecology and management**, v. 411, p. 35-45. 2018.

PALMA, R. A. Determination of aboveground carbon density of mangium (*Acacia mangium* Willd.) using biomass expansion factor. **Mindanao Journal of Science and Technology**, v.12, p. 39-50, 2014.

PÁSCOA, F.; MARTINS, F.; GONZÁLES, J. C. R. S.; Estabelecimento simultâneo de equações de biomassa para o pinheiro bravo. **Anais do 2º Simpósio Iberoamericano de Gestion y Economía Forestal**. p. i-f, 2004.

PAYN, T.; CARNUS, J. M.; FREER-SMITH, P.; KIMBERLEY, M.; KOLLERT, W.; LIU, S.; WINGFIELD, M. J. Changes in planted forests and future global implications. **Forest Ecology and Management**, v. 352, p. 57-67, 2015.

PEERBHAY, K. Y.; MUTANGA, O.; ISMAIL, R. Random Forests Unsupervised Classification: The Detection and Mapping of *Solanum mauritianum* Infestations in Plantation Forestry Using Hyperspectral Data. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v. 8, n. 6, p. 3107-3122. 2015.

PÉLLICO NETTO, S.; BEHLING, A. Additivity of tree biomass components using ratio estimate. **Annals of the Brazilian Academy of Sciences**. 31p. 2019.

PHILLIPS, O. L.; SULLIVAN, M. J.; BAKER, T. R.; MENDOZA, A. M.; VARGAS, P. N.; VÁSQUEZ, R. Species Matter: Wood Density Influences Tropical Forest Biomass at Multiple Scales. **Surveys in Geophysics**, p. 1-23. 2019.

POORTER, H.; NIKLAS, K.J.; REICH, P.B.; OLCKSYN, J.; POOT, P.; MOMMER, L. Biomass allocation to leaves, stems and roots: meta-analyses of interspecific variation and environmental control. **New Phytologist**, v. 193, p. 30-50, 2012

RATUCHNE, L. C.; KOEHLER, H. S.; WATZLAWICK, L. F.; SANQUETTA, C. R.; SCHAMNE, P. A. Estado da arte na quantificação de biomassa em raízes de formações florestais. **Floresta e Ambiente**, v. 23, n. 3, p. 450-462. 2016.

RÄTY, J.; PACKALEN, P.; MALTAMO, M. Nearest neighbor imputation of logwood volumes using bi-temporal ALS, multispectral ALS and aerial images. **Scandinavian Journal of Forest Research**, p. 1-15. 2019.

REGO, A. M.; DA SILVA, M. D. M.; SOBRINHO, O. P. L.; FERREIRA, J. C. S. Modelos de equações volumétricas para um povoamento de eucalipto situado em Codó (MA). **Revista Ibero-Americana de Ciências Ambientais**, v. 10, n. 4, p. 1-8. 2019.

REZENDE, A. V.; VALE, A. T.; SANQUETTA, C. R.; FIGUEIREDO FILHO, A.; FELFILI, J. M. Comparação de modelos matemáticos para estimativa do volume, biomassa e estoque de carbono da vegetação lenhosa de um cerrado sensu stricto em Brasília, DF. **Scientia Florestalis**, n. 71, p. 65-76, 2006.

RFF– RESOURCES FOR THE FUTURE. The Social Cost of Carbon Initiative. Link: <<https://www.rff.org/social-cost-carbon-initiative/>> Acesso: 12/01/2020.

ROBORTELLA, H. S. Equações de biomassa e estoques de carbono de seis espécies em plantios mistos no noroeste do Mato Grosso. 2010.

ROLIM, S. G.; COUTO, H. T. Z. D.; JESUS, R. M. D.; FRANÇA, J. T. Modelos volumétricos para a floresta nacional do Tapirapé-aquirí, Serra dos Carajás (PA). **Acta Amazonica**, v. 36, n. 1, p. 107-114. 2006

RUIVO, L. V.; STONE, A. I.; FIENUP, M. Reproductive status affects the feeding ecology and social association patterns of female squirrel monkeys (*Saimiri collinsi*) in an Amazonian rainforest. **American journal of primatology**, v. 79, n. 6, p. 22657. 2017.

SAINT-ANDRÉ, L.; M'BOU, A.T.; MABIALA, A.; MOUVONDY, W.; JOURDAN, C.; ROUPSARD, O.; DELEPORTE, P.; HAMEL, O.; NOUVELLON, Y. Age-related equations for above – and below – ground biomass of a *Eucalyptus* hybrid in Congo. **Forest Ecology and Management**, v. 205, p. 199-214, 2005.

SAKICI, O. E.; GÜNLÜ, A. Artificial intelligence applications for predicting some stand attributes using Landsat 8 OLI satellite data: A case study from Turkey. **Applied Ecology and Environmental Research**, v. 16, n. 4, p. 5269-5285. 2018.

SALCEANU, A. JULIA 1.0 by example. **Packt Publishing**. 378 p., 2018a.

SALCEANU, A. Julia Programming Projects. **Packt Publishing**, 500 p., 2018b.

SANQUETTA, C.R.; CORTE, A.P.D.; SILVA, F. da. Biomass expansion factor and root-to-shoot ratio for *Pinus* in Brazil. **Carbon Balance and Management**, v. 6, n. 6, p. 1-8, 2011.

SANQUETTA, C. R.; DALLA CORTE, A. P.; PELISSARI, A. L.; TOMÉ, M.; MAAS, G. C. B.; SANQUETTA, M. N. I. Dinâmica em superfície, volume, biomassa e carbono nas florestas plantadas brasileiras: 1990-2016. **Biofix Scientific Journal**, v. 3, n. 1, p. 152-160. 2018a.

SANQUETTA, C. R.; DOLCI, M.; CORTE, A. P. D.; SANQUETTA, M. N. I.; PELISSARI, A. L. Estimação de volumes de *Araucaria angustifolia* (Bertol.) O. Kuntze por fatores de forma em classes diamétricas e modelos de regressão. **Centro Científico Conhecer**, v. 13, n. 23, p. 588-597, 2016a.

SANQUETTA, M. N. I.; COUTINHO, V. M.; BEHLING, A.; CORTE, A. P. D.; SANQUETTA, C. R. O uso de modelos alométricos tradicionais na estimativa do peso seco aéreo individual para acácia-negra. **Enciclopédia Biosfera**, v. 12, n. 23, p. 842-854, 2016b.

SANQUETTA, C. R.; PIVA, L. R.; WOJCIECHOWSKI, J.; CORTE, A. P. D.; SCHIKOWSKI, A. B. Volume estimation of *Cryptomeria japonica* logs in southern Brazil using artificial intelligence models. **Southern Forests: a Journal of Forest Science**, v. 80, n. 1, p. 29-36, 2018b.

SANQUETTA, C. R.; WATZLAWICK, L. F.; ARCE, J. E.; DE MELLO, A. A. Ecuaciones de biomasa aérea y subterránea en plantaciones de *Pinus taeda* en el sur del estado de Paraná, Brasil. In **Simpósio Internacional Medición y Monitoreo de la Captura de Carbono en Ecosistemas Forestales**. 2001.

SANQUETTA, C. R. WATZLAWICK, L. F.; BALBINOT, R.; ZILOTTO, M. A.; GOMES, F. D. As florestas e o carbono. **Imprensa Universitaria Universidade Federal do Paraná**. p. 1-140, 2002.

SANQUETTA, C. R.; BEHLING, A.; DALLA CORTE, A. P.; SIMON, A.; PSCHIEDT, H.; RUZA, M. S.; MOCHIUTTI, S. Estoques de biomassa e carbon em povoamentos de acácia negra em diferentes idades no Rio Grande do Sul. **Scientia Forestalis**, v. 42, n. 103, p. 361-370. Setembro, 2014b.

SANQUETTA, C. R.; JUNIOR, S. C.; SANQUETTA, M. N. I.; DALLA CORTE, A. P.; PEDROSA-MACEDO, J. H. PARTIÇÃO DA BIOMASSA E RELAÇÕES BIOMÉTRICAS DE AROEIRA-PIMENTEIRA. **BIOFIX Scientific Journal**, v. 4, n. 2, p. 104-109. 2019.

SANQUETTA, C.R.; WATZLAWICK, L. F.; CORTE, A.P.D.; FERNANDES, L. A. V.; SIQUEIRA, J. D. P. Inventários Florestais: Planejamento e Execução - 3ª EDIÇÃO. **Multi-graphic**. 406p., 2014a.

SANQUETTA, C. R.; DALLA CORTE, A. P.; MOGNON, F.; MAAS, G. C. B.; RODRIGUES, A. L. Estimativa de carbono individual para *Araucaria angustifolia*. *Pesquisa Agropecuária Tropical*, v.44, n.1, p.1-8, 2014c.

SANQUETTA, C. R.; WOJCIECHOWSKI, J.; DALLA CORTE, A. P.; BEHLING, A.; NETTO, S. P.; RODRIGUES, A. L.; SANQUETTA, M. N. Comparison for data mining and allometric model in estimation of tree biomass. **BMC Bioinformatics**, v. 16, n. 1, p. 247-256, 2015.  
Link:<<https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-015-0662-5>> Acesso: 04/01/2020.

SANTOS, F. G.; CAMARGO, P. B.; OLIVEIRA, R. C. Estoque e dinâmica de biomassa arbórea em floresta ombrófila densa na flona tapajós: Amazônia oriental. **Ciência Florestal**, v. 28, n. 3, p. 1049-1059. 2018.

SCIKITLEARN. sklearn.neural\_network.MLPRegressor. 2020a. Link: <[https://scikit-learn.org/stable/modules/generated/sklearn.neural\\_network.MLPRegressor.html#sklearn.neural\\_network.MLPRegressor](https://scikit-learn.org/stable/modules/generated/sklearn.neural_network.MLPRegressor.html#sklearn.neural_network.MLPRegressor)>

SCIKITLEARN. sklearn.neighbors.KNeighborsRegressor. 2020b. Link: <<https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsRegressor.html#sklearn.neighbors.KNeighborsRegressor>>.

SCIKITLEARN. sklearn.ensemble.RandomForestRegressor. 2020c. Link: <<https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html#sklearn.ensemble.RandomForestRegressor>>

SCIKITLEARN. sklearn.svm.SVR. 2020d. Link: <<https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVR.html#sklearn.svm.SVR>>

SCHILOWSKI, A. B.; CORTE, A. P. D.; RUZA, M. S.; SANQUETTA, C. R.; MONTAÑO, R. A. Modeling of stem form and volume through machine learning. **Anais da Academia Brasileira de Ciências**, v. 90, n. 4, p. 3389-3401. 2018.

SCHILOWSKI, A. B.; DALLA CORTE, A. P.; SANQUETTA, C. R.; RODRIGUES, L.; VASCONCELLOS, B. N. Análise e ajuste do fator de expansão de biomassa e razão raízes-parte aérea para Álamo. **Enciclopédia Biosfera**, v. 11, n. 21, p. 107-119, 2015.

SCHNEIDER, P. R.; FINGER, C. A. G.; SOBRINHO, V. G.; SCHNEIDER, P. S. P. Determinação indireta do estoque de biomassa e carbono em povoamentos de acácia-negra (*Acacia mearnsii* De Wild.). **Ciência Florestal**, v. 15, n. 4, p. 391-402, 2005.

SCHUH, M. S. Modelos preditivos de biomassa em Floresta Amazônica a partir de dados lidar. Dissertação de Mestrado curso de Engenharia Florestal da UFSM, 81p, 2019.

SENGUPTA, A. Julia High performance. **Packt Publishing**, 132 p., 2016.

SENGUPTA, A. Julia High performance. **Packt Publishing**, 218 p., 2018.

SHARIFI, A.; AMINI, J.; TATEISHI, R. Estimation of forest biomass using multivariate relevance vector regression. **Photogrammetric Engineering & Remote Sensing**, v. 82, n. 1, p. 41-49. 2016.

SHARIF, U.; MEHMOOD, Z.; MAHMOOD, T.; JAVID, M. A.; REHMAD, A.; SABA, T. Scene analysis and search using local features and support vector machine for effective content-based image retrieval. **Artificial Intelligence Review**, v. 52, n. 2, p. 901-925. 2019.

SHRESTHA, N. K.; SHUKLA, S. Support vector machine based modeling of evapotranspiration using hydro-climatic variables in a sub-tropical environment. **Agricultural and Forest Meteorology**, v. 200, p. 172-184, 2015.

SHERRINGTON, M. Mastering Julia - Tackle the Contemporary Challenges of Programming and Data Science with Julia. **Packt Publishing**. 412 p. 2015.

SHERRINGTON, M. Mastering Julia - Solve complex data processing problems with Julia. **Packt Publishing**, 463 p., 2019.

SHERRY, T.; HOLMES, D.; SICCAM, T. G. Forest Inventory of a Northern Hardwood Forest: Bird Area at the Hubbard Brook Experimental Forest, p. 1981. 2019.

SILVA, C.; KLAUBERD, C.; HUDAK, A.; VIERLING, L.; JAAFAR, W.; MOHAN, M.; SAATCHI, S. Predicting stem total and assortment volumes in an industrial Pinus taeda L. forest plantation using airborne laser scanning data and random forest. **Forests**, v. 8, n. 7, p. 254-271. 2017.

SILVA, C. A.; KLAUBERG, C.; HUDAK, A. T.; VIERLING, L. A.; LIESENBERG, V.; BERNETT, L. G.; SCHOENINGER, E. R. Estimating Stand Height and Tree Density in Pinus taeda plantations using in-situ data, airborne LiDAR and k-Nearest Neighbor Imputation. **Anais da Academia Brasileira de Ciências**, v. 90, n. 1, p. 295-309. 2018.

SILVA, L. F.; FERREIRA, G. L.; SANTOS, A. C. A.; LEITE, H. G.; SILVA, M. L. Equações hipsométricas, volumétricas e de crescimento para Khaya ivorensis plantada em Pirapora. **Floresta e Ambiente**, v. 23, n. 3, p. 362-368. 2016.

SILVASANKAR, T.; LONE, J. M.; SARMA, K. K.; QADIR, A.; RAJU, P. L. N. Estimation of Above Ground Biomass Using Support Vector Machines and ALOS/PALSAR data. **Vietnam Journal of Earth Sciences**, v. 41, n. 2, p. 95-104. 2019.

SILVEIRA, P. Ajuste de modelos matemáticos para estimar biomassa aérea em Floresta Ombrófila Densa. **Floresta, Curitiba**, v. 39, n. 4, p. 743-752, 2009.

SOUZA, D. V.; NIEVOLA, J. C.; SANTOS, J. X.; WOJCIECHOWSKI, J.; GONÇALVES, A. L.; CORTE, A. P. D.; SANQUETTA, C. R. k-Nearest Neighbor Regression in the Estimation of Tectona G randis Trunk Volume in the State of Pará, Brazil. **Journal of Sustainable Forestry**, p. 1-14. 2019.

SUKHODOLOV, A. P.; SOROKINA, P. G.; FEDOTOV, A. P. Numerical analysis of ecology-economic models for forest fire fighting in Baikal region. **Discrete & Continous Models & Applied Computational Science**. v. 27, n. 2, p. 154-164, 2019.

SUTTON, R. S.; BARTO, A. G. Reinforcement learning: An introduction. Massachusetts: MIT Press. 1998.

TABARI, H.; KISI, O.; EZANI, A. ; TALAEI, P. H.; SVM, ANFIS, regression and climate based models for reference evapotranspiration modeling using limited climatic data in a semi-arid highland environment. **Journal of Hydrology**, v. 444, p. 78-89, 2012.

TAVARES JÚNIOR, I. D. S.; ROCHA, J. E. C. D.; EBLING, Â. A.; CHAVES, A. D. S.; ZANUNCIO, J. C.; FARIAS, A. A.; LEITE, H. G. Artificial Neural Networks and Linear Regression Reduce Sample Intensity to Predict the Commercial Volume of Eucalyptus Clones. **Forests**, v. 10, n. 3, p. 268. 2019.

THOMAS, D. C.; VANDEGRIFT, R.; LUDDEN, A.; CARROLL, G. C.; ROY, B. A. Spatial ecology of the fungal genus *Xylaria* in a tropical cloud forest. **Biotropica**, v. 48, n. 3, p. 381-393. 2016.

VAHEDI, A. A. Artificial neural network application in comparison with modeling allometric equations for predicting above-ground biomass in the Hyrcanian mixed-beech forests of Iran. **Biomass and Bioenergy**, v. 88, p. 66-76. 2016.

VALERIO, A. F.; WATZLAWICK, L. F.; DOS SANTOS, R. T.; SILVESTRE, R.; KOEHLER, H. S. Ajuste de modelos matemáticos para estimativa de biomassa nos distintos componentes de uma do Japão (*Hovenia dulcis* Thunb.). **AMBIÊNCIA**, v.3, n. 3, p. 325-336, 2007.

VAN BICH, N.; EYLES, A.; MENDHAM, D.; DONG, T.; RATKOWSKY, D.; EVANS, K.; MOHAMMED, C. Contribution of harvest residues to nutrient cycling in a tropical *Acacia mangium* Willd. plantation. **Forests**, v. 9, n. 9, p. 577. 2018.

VAPNIK, V. N. The nature of statistical learning theory. **Springer science & business media**, 2013.

VENDRUSCOLO, D. G. S.; DRESCHER, R.; SOUZA, H. S.; MOURA, J. P. V. M.; MAMORÉ, F. M. D.; SIQUEIRA, T. D. S. Estimativa da altura de eucalipto por meio de regressão não linear e redes neurais artificiais. **Revista Brasileira de Biometria**, v. 33, n. 4, p. 556-569, 2015.

VIEIRA, S. A.; ALVES, L. F.; AIDAR, M. P. M.; ARAÚJO, L. S.; BAKER, T.; BATISTA, J. L. F.; CAMPOS, M. C. R.; CAMARGO, P. B.; CHAVE, J.; DELITTI, W. B.; HIGUCHI, N.; HONÓRIO, E.; JOLY, C. A.; KELLER, M.; MARTINELLI, L. A.; MATTOS, E. A.; METZKER, T.; PHILLIPS, O.; SANTOS, F. A. M.; SHIMABUKURO, M. T.; SILVEIRA, M.; TRUMBORE, S. E. Estimation of biomass and carbon stocks: the case of the Atlantic Forest. **Biota Neotropica**, v. 8, n. 2. 2008.

VOULGARIS, Z. Julia for Data Science. **Technics Publications**. 366 p. 2016.

WANG, J.; XU, J.; PENG, Y.; WANG, H.; SHEN, J. Prediction of forest unit volume based on hybrid feature selection and ensemble learning. **Evolutionary Intelligence**, p. 1-12. 2019.

WANG, L. Support vector machines: theory and applications. **Springer Science & Business Media**, v.177, 46p. 2005.

WALTHER, B. A.; MOORE, J. L. The concepts of bias, precision and accuracy, and their use in testing the performance of species richness estimators, with a literature review of estimator performance. **ECOGRAPHY**, v. 28, p.815-829, 2005.

WATSON, G. L.; TELESKA, D.; REID, C. E.; PFISTER, G. G.; JERRETT, M. Machine learning models accurately predict ozone exposure during wildfire events. **Environmental Pollution**, v. 254, p. 112792. 2019.



WERBOS, P. Backpropagation Through Time: What It Does and How to Do It. **Proceedings of the IEEE**.v. 78, n.10, p. 1550–1560, 1990.

WILKINS, D. C.; BUCHANAN, B. G. **On debugging rule sets when reasoning under uncertainty**. Knowledge Systems Laboratory, Department of Computer Science, Stanford University, 1987.

XING, J.; LUO, K.; WANG, H.; FAN, J. Estimating biomass major chemical constituents from ultimate analysis using a random forest model. **Bioresource technology**, v. 288, p. 121541. 2019.

YOUNG, W. S.; GIL, J. C.; UK, K. J.; JOON, K. S. Assessment of climate change impact on aquatic ecology health indices in Han river basin using SWAT and random forest. **Journal of Korea Water Resources Association**, v. 51, n. 10, p. 863-874. 2018.

YURET, D. Knet: beginning deep learning with 100 lines of Julia. **30th Conference on Neural Information Processing Systems**, 7p. 2016.

ZAMAR, D. S.; GOPALUNI, B.; SOKHANSANJ, S. A Constrained K-Means and Nearest Neighbor Approach for Route Optimization in the Bale Collection Problem. **IFAC-PapersOnLine**, v. 50, n. 1, p. 12125-12130. 2017.

ZAR, J. H. Biostatistical analysis. **Prentice Hall**. v. 5, 944p. 2009.

ZHANG, D.; STENGER, A.; HAROU, P. A. Policy instruments for developing planted forests: theory and practices in China, the U.S., Brazil, and France. **Journal of Forest Economics**, v. 21, n. 4, p. 223-237, 2015.

ZHANG-JAMES, Y.; GLATT, S. J.; FARAONE, S. V. Nu Support Vector Machine in Prediction of Fluid Intelligence Using MRI Data. In Challenge in Adolescent Brain Cognitive Development Neurocognitive Prediction. **Springer**. p. 92-98. 2019.

ZHAO, D.; KANE, M.; MARKEWITZ, D.; TESKEY, R.; CLUTTER, M. Additive tree biomass equations for mid-rotation loblolly pine plantations. **Forest Science**, v. 61, n. 4, p. 613-623, 2015

ZHU, X.; LIU, D. Improving forest aboveground biomass estimation using seasonal Landsat NDVI time-series. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 102, p. 222-231. 2015.